

- ANDERSON, N.L. AND ANDERSON, N.G. (1991). A two-dimensional gel database of human plasma proteins. *Electrophoresis*, **12**, 883-906.
- ANDERSON, N.L., ESQUER-BLASCO, R., HOFMANN, J.P. AND ANDERSON, N.G. (1991). A two-dimensional gel database of rat liver proteins useful in gene regulation and drug effects studies. *Electrophoresis*, **12**, 907-930.
- ANDERSON, N.L., COPPLE, D.C., BENDELE, R.A., PROBST, G.S., RICHARDSON, F.C. (1992). Covalent protein modifications and gene expression changes in rodent liver following administration of methapyrilene: a study using two-dimensional electrophoresis. *Fundamental and Applied Toxicology*, **18**, 570-580.
- APPEL, R.D., BAIROCH, A. AND HOCHSTRASSER, D.F. (1994). A new generation of information retrieval tools for biologists: the example of the ExPASy WWW server. *Trends in Biochemical Sciences*, **19**, 258-260.
- APPEL, R.D., HOCHSTRASSER, D.F., FUNK, M., VARGAS, J.R., PELIGRINI, C., MÜLLER, A.F. AND SCHERRER, J.R. (1991). The MELANIE project: from a biopsy to automatic protein map interpretation by computer. *Electrophoresis*, **12**, 722-735.
- APPEL, R.D., SANCHEZ, J.-C., BAIROCH, A., GOLAZ, O., MIU, M., VARGAS, J.R. AND HOCHSTRASSER, D.F. (1993). SWISS-2DPAGE: a database of two-dimensional gel electrophoresis images. *Electrophoresis*, **14**, 1323-1328.
- APPEL, R.D., SANCHEZ, J.-C., BAIROCH, A., GOLAZ, O., RAVIER, F., PASQUALI, C., HUGHES, G. AND HOCHSTRASSER, D.F. (1994). The SWISS-2DPAGE database of two-dimensional polyacrylamide gel electrophoresis. *Nucleic Acids Research*, **22**, 3581-3582.
- BAIROCH, A. AND BOECKMANN, B. (1994). The SWISS-PROT protein sequence databank: current status. *Nucleic Acids Research*, **22**, 3578-3580.
- BAKER, C.S., CORBETT, J.M., MAY, A.J., YACOB, M.H. AND DUNN, M.J. (1992). A human myocardial two-dimensional electrophoresis database: protein characterisation by microsequencing and immunoblotting. *Electrophoresis*, **13**, 723-726.
- BARTLET-JONES, M., JEFFERY, W.A., HANSEN, H.F. AND PAPPIN, D.J.C. (1994). Peptide ladder sequencing by mass spectrometry using a novel, volatile degradation reagent. *Rapid Communications in Mass Spectrometry*, **8**, 737-742.
- BAUER, D., MÜLLER, H., REICH, J., RIEDEL, H., AHRENKIEL, V., WARTHOF, P. AND STRAUSS, M. (1993). Identification of differentially expressed mRNA species by an improved display technique (DDRT-PCR). *Nucleic Acids Research*, **21**, 4272-4280.
- BERNERS-LEE, T.J., CAILIAU, R., GROFF, J.F. AND POLLERMAN, B. (1992). *Electronic Networking: Research, Applications, and Policy*, **2**, 52-58.
- BIRD, A.P. (1995). Gene number, noise reduction and biological complexity. *Trends in Genetics*, **11**, 94-100.
- BJELLOVIST, B., EK, K., RIGHETTI, P.G., GIANAZZA, E., GORG, A., WESTERMEIER, R. AND POSTEL, W. (1982). Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications. *Journal of Biochemical and Biophysical Methods*, **6**, 317-339.
- BJELLOVIST, B., PASQUALI, C., RAVIER, F., SANCHEZ, J.-C. AND HOCHSTRASSER, D.F. (1993a). A nonlinear wide-range immobilized pH gradient for two-dimensional electrophoresis and its definition in a relevant pH scale. *Electrophoresis*, **14**, 1357-1365.
- BJELLOVIST, B., SANCHEZ, J.-C., PASQUALI, C., RAVIER, F., PAQUET, N., FRUTIGER, S., HUGHES, G.J. AND HOCHSTRASSER, D.F. (1993b). Micropreparative 2-D electrophoresis allowing the separation of milligram amounts of proteins. *Electrophoresis*, **14**, 1375-1378.
- BJELLOVIST, B., HUGHES, G., PASQUALI, C., PAQUET, N., RAVIER, F., SANCHEZ, J.-C., FRUTIGER, S. AND HOCHSTRASSER, D. (1993c). The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis*, **14**, 1023-1031.
- BONNER, W.M. AND LASKEY, R.A. (1974). A film detection method for tritium-labeled proteins and nucleic acids in polyacrylamide gels. *European Journal of Biochemistry*, **46**, 83-88.
- BOITTELL, T., GARRELS, J.I., FRANZA, B.R., MONARDO, P.J. AND LATTER, G.I. (1994). REF52 on global gel navigator: an internet-accessible two-dimensional gel electrophoresis database. *Electrophoresis*, **15**, 1487-1490.
- BREWER, J., GRUND, E., HAGERLID, P., OLSSON, I. AND LIZANA, J. (1986). In *Electrophoresis '86* (M.J. Dunn, Ed.), pp. 226-229. VCH, Weinheim.

- CARR, S.A., HUDDLESTON, M.J. AND BEAN, M.F. (1993). Selective identification and differentiation of N- and O-linked oligosaccharides in glycoproteins by liquid chromatography-mass spectrometry. *Protein Science*, **2**, 183-196.
- CAMPBELL, K.P., MACLENNAN, D.H. AND JORGENSEN, A.O. (1983). Staining of the Ca²⁺-binding proteins, calsequestrin, calmodulin, troponin C, and S-100, with the cationic dye 'Stains-all'. *Journal of Biological Chemistry*, **258**, 11267-11273.
- CELIS, J.E., CRUGER, D., KIL, J., DEJGAARD, K., LAURIDSEN, J.B., RATZ, G.P., BASSE, B., CELIS, A., RASMUSSEN, H.H., BAUW, G. AND VANDEKERKHOVE, J. (1990a). A two-dimensional gel protein database of noncultured total normal human epidermal keratinocytes: identification of proteins strongly up-regulated in psoriatic epidermis. *Electrophoresis*, **11**, 242-254.
- CELIS, J.E., GESSER, B., RASMUSSEN, H.H., MADSEN, P., LEFFERS, H., DEJGAARD, K., HONORE, B., OLSEN, E., RATZ, G., LAURIDSEN, J.B., BASSE, B., MOURIZTEN, S., HELLERUP, M., ANDERSEN, A., WALBUM, E., CELIS, A., BAUW, G., PUYPE, M., VAN DAMME, J. AND VANDEKERKHOVE, J. (1990b). Comprehensive two-dimensional gel protein database offers a global approach to the analysis of human cells: the transformed amnion cells (AMA) master database and its link to genome DNA sequence data. *Electrophoresis*, **11**, 898-1071.
- CELIS, J.E., RASMUSSEN, H.H., OLSEN, E., MADSEN, P., LEFFERS, H., HONORE, B., DEJGAARD, K., GROMOV, P., HOFFMANN, H.J., NIELSEN, M., VASSILEV, A., VINTERMYR, O., HAO, J., CELIS, A., BASSE, B., LAURIDSEN, J., RATZ, G.P., ANDERSEN, A.H., WALBUM, E., KJAERGAARD, I., PUYPE, M., VAN DAMME, J. AND VANDEKERKHOVE, J. (1993). The human keratinocyte two-dimensional database: update 1993. *Electrophoresis*, **14**, 1091-1198.
- CELIS, J.E. AND OLSEN, E. (1994). A qualitative and quantitative protein database approach identified individual and groups of functionally related proteins that are differentially regulated in simian virus 40 (SV40) transformed human keratinocytes: an overview of the functional changes associated with the transformed phenotype. *Electrophoresis*, **15**, 309-344.
- CORBETT, J.M., DUNN, M.J., POSCH, A. AND GORG, A. (1994a). Positional reproducibility of protein spots in two-dimensional polyacrylamide electrophoresis using immobilized pH gradient isoelectric focusing in the first dimension - an interlaboratory comparison. *Electrophoresis*, **15**, 1205-1211.
- CORBETT, J.M., WHEELER, C.H., BAKER, C.S., YACOUR, M.H. AND DUNN, M.J. (1994b). The human myocardial two-dimensional gel protein database: update 1994. *Electrophoresis*, **15**, 1459-1465.
- CORDWELL, S., WILKINS, M.R., CERPA-POLJAK, A., GOOLEY, A.A., DUNCAN, M., WILLIAMS, K.L. AND HUMPHERY-SMITH, I. (1995). Cross-species identification of proteins separated by two-dimensional electrophoresis using MALDI-TOF and amino acid composition. *Electrophoresis*, **15**, 438-443.
- COTTRELL, J.S. (1994). Protein identification by peptide mass fingerprinting. *Peptide Research*, **7**, 115-124.
- CRIMMINS, D.L., MCCOURT, D.W., THOMA, R.S., SCOTT, M.G., MACKE, K. AND SCHWARTZ, B.D. (1990). *In situ* chemical cleavage of proteins immobilized to glass-fiber and polyvinylidene fluoride membranes: cleavage at tryptophan residues with 2-(2'-nitrophenylsulfonyl)-3-methyl-3'-bromoindolenine to obtain internal amino acid sequence. *Analytical Biochemistry*, **187**, 27-38.
- DUTHEL, S. AND REVOL, A. (1993). Glycan microheterogeneity of alpha 1-antitrypsin in serum and meconium from normal and cystic fibrosis patients by crossed immunoaffinoelectrophoresis with different lectins (Con A, LCA, WGA). *Clinical and Chemical Acta*, **215**, 173-187.
- ECKERSKORN, C., JUNGBLUT, P., MEWES, W., KLOSE, J. AND LOTTSPEICH, F. (1988). Identification of mouse brain proteins after two-dimensional electrophoresis and electroblotting by microsequence analysis and amino acid composition. *Electrophoresis*, **9**, 830-838.
- ECKERSKORN, C., STRUPAT, K., KARAS, M., HILLENKAMP, F. AND LOTTSPEICH, F. (1992).

- Mass spectrometric analysis of blotted proteins after gel electrophoretic separation by matrix-assisted laser desorption/ionization. *Electrophoresis*, **13**, 664-665.
- ECKERSKORN, C. AND LOTTSCHEICH, F. (1993). Structural characterisation of blotting membranes and the influence of membrane parameters for electroblotting and subsequent amino acid sequence analysis of proteins. *Electrophoresis*, **14**, 831-838.
- EK, K., BJELLQVIST, B.J. AND RIGHETTI, P.G. (1983). Preparative isoelectric focusing in immobilized pH gradients. I. General principles and methodology. *Journal of Biochemical and Biophysical Methods*, **8**, 135-155.
- FEY, S.J., CARLSEN, J., MOSE LARSEN, P., JENSEN, U.A., KJELDSEN, K. AND HAUNSO, S. (1994). Two-dimensional gel electrophoresis as a tool for molecular cardiology. Proceedings of the International Society for Heart Research 'XV European Section Meeting', pp 9-16.
- FREY, J.R., KUHN, L., KETTMAN, J.R. AND LEFKOVITS, I. (1994). The amino acid composition of 350 lymphocyte proteins. *Molecular Immunology*, **31**, 1219-1231.
- GARRELS, J.I. (1989). The QUEST system for quantitative analysis of two-dimensional gels. *Journal of Biological Chemistry*, **264**, 5269-5282.
- GARRELS, J.I. AND FRANZA, B.R. (1989). The REF52 protein database. *Journal of Biological Chemistry*, **264**, 5283-5298.
- GARRELS, J.I., FRANZA, B.R., CHANG, C. AND LATTER, G. (1990). Quantitative exploration of the REF52 protein database: cluster analysis reveals the major protein expression profiles in responses to growth regulation, serum stimulation, and viral transformation. *Electrophoresis*, **11**, 1114-1130.
- GARRELS, J.I., FUTCHER, B., KOBAYASHI, R., LATTER, I., SCHWENDER, B., VOLPE, T., WARNER, J.R. AND MCLAUGHLIN, C.S. (1994). Protein identification for a *Saccharomyces cerevisiae* protein database. *Electrophoresis*, **15**, 1466-1486.
- GELFI, C., BOSSI, M.L., BJELLQVIST, B. AND RIGHETTI, P.G. (1987). Isoelectric focusing in immobilized pH gradients in the pH 10-11 range. *Journal of Biochemical and Biophysical Research Methods*, **15**, 41-48.
- GHARAHDAAGHI, F., ATHERTON, D., DEMOTT, M. AND MISCHKE, S.M. (1992). Amino acid analysis of PVDF-bound proteins. in *Techniques in Protein Chemistry III* (R.H. Ageletti, Ed.), pp 249-260. Academic Press, San Diego.
- GHOSH, P., OKOH, C., LIU, Q.H. AND LAKSHMAN, M.R. (1993). Effects of chronic ethanol on enzymes regulating sialylation and desialylation of transferrin in rats. *Alcoholism: Clinical and Experimental Research*, **17**, 576-579.
- GIOMETTI, C.S., GEMMELL, M.A., TOLLAKSEN, S.L. AND TAYLOR, J. (1991). Quantitation of human leukocyte proteins after silver staining: a study with two-dimensional electrophoresis. *Electrophoresis*, **12**, 536-543.
- GIOMETTI, C.S., TAYLOR, J. AND TOLLAKSEN, S.L. (1992). Mouse liver protein database: a catalog of proteins detected by two-dimensional gel electrophoresis. *Electrophoresis*, **13**, 970-991.
- GOLAZ, O., HUGHES, G.J., FRUTIGER, S., PAQUET, N., BAIROCH, A., PASQUALI, C., SANCHEZ, J.C., TISSOT, J.D., APPEL, R.D., WALZER, C., BALANT, L. AND HOCHSTRASSER, D.F. (1993). Plasma and red blood cell protein maps: update 1993. *Electrophoresis*, **14**, 1223-1231.
- GOLD, M.R., YUNGWIRTH, T., SUTHERLAND, C.L., INGHAM, R.J., VIANZON, D., CHIU, R., VAN-OOSTVEEN, I., MORRISON, H.D. AND AEBERSOLD, R. (1994). Purification and identification of tyrosine-phosphorylated proteins from lymphocytes stimulated through the antigen receptor. *Electrophoresis*, **15**, 441-453.
- GOLDBERG, H.A., DOMENICUCCI, C., PRINGLE, G.A. AND SODEK, J. (1988). Mineral-binding proteoglycans of fetal porcine calvarial bone. *Journal of Biological Chemistry*, **263**, 12092-12101.
- GOOLEY, A.A., MARSHCHALEK, R. AND WILLIAMS, K.L. (1992). Size polymorphisms due to changes in the number of O-glycosylated tandem repeats in the *Dictyostelium discoideum* glycoprotein PnA. *Genetics*, **130**, 749-756.
- GORG, A., POSTEL, W. AND GUNTHER, S. (1988). The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis*, **9**, 531-546.
- GORG, A., POSTEL, W., GUNTHER, S., WESER, J., STRAILER, J.R., HANASHI, S.M., SOMERLOT, L.

- AND KUICK, R. (1988). Approach to stationary two-dimensional pattern: influence of focusing time and immobilize/carrier ampholyte concentrations. *Electrophoresis*, **9**, 37-46.
- GRAVEL, P., GOLAZ, O., WALZER, C., HOCHSTRASSER, D.F., TURLER, H., AND BALANT, L.P. (1994). Analysis of glycoproteins separated by two-dimensional gel electrophoresis using lectin blotting revealed by chemiluminescence. *Analytical Biochemistry*, **221**, 66-71.
- GUNTHER, S., POSTEL, W., WIERING, H. AND GORG, A. (1988). Acid phosphatase typing for breeding nematode-resistant tomatoes by isoelectric focusing with an ultranarrow immobilized pH gradient. *Electrophoresis*, **9**, 618-620.
- HANASH, S.M., STRAHLER, J.R., NEEL, J.V., HAILAT, N., MELHEM, R., KEIM, D., ZHU, X.X., WAGNER, D., GAGE, D.A. AND WATSON, J.T. (1991). Highly resolving two-dimensional gels for protein sequencing. *Proceedings of the National Academy of Sciences USA*, **88**, 5709-5713.
- HARRINGTON, M.G., COFFMAN, J.A., CALZONE, F.J., HOOD, L.E., BRITTEN, R.J. AND DAVIDSON, E.H. (1992). Complexity of sea urchin embryo nuclear proteins that contain basic domains. *Proceedings of the National Academy of Sciences USA*, **89**, 6252-6256.
- HARRINGTON, M.G., LEE, K.H., YUN, M., ZEWEIT, T., BAILEY, J.E. AND HOOD, L.E. (1993). Mechanical precision in two-dimensional electrophoresis can improve spot positional reproducibility. *Applied and Theoretical Electrophoresis*, **3**, 347-353.
- HELLMAN, U., WERNSTEDT, C., GONEZ, J. AND HELDIN, C.-H. (1995). Improvement of an in-gel digestion for the micropreparation of internal protein fragments for amino acid sequencing. *Analytical Biochemistry*, **224**, 451-455.
- HENZEL, W.J., BILLECI, T.M., STULTS, J.T., WONG, S.C., GRIMLEY, C. AND WATANABE, C. (1993). Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proceedings of the National Academy of Sciences USA*, **90**, 5011-5015.
- HESS, D., COVEY, T.C., WINZ, R., BROWNSEY, R.W. AND AEBERSOLD, R. (1993). Analytical and micropreparative peptide mapping by high performance liquid chromatography/electrospray mass spectrometry of proteins purified by gel electrophoresis. *Protein Science*, **2**, 1342-1351.
- HOBOM, U., HOUTHAËVE, T. AND SANDER, C. (1994). Amino acid analysis and protein database compositional search as a rapid and inexpensive method to identify proteins. *Analytical Biochemistry*, **222**, 202-209.
- HOCHSTRASSER, D.F. AND MERRIL, C.R. (1988). 'Catalysts' for polyacrylamide gel polymerization and detection of proteins by silver staining. *Applied and Theoretical Electrophoresis*, **1**, 35-40.
- HOCHSTRASSER, D.F., PATCHORNIK, A. AND MERRIL, C.R. (1988). Development of polyacrylamide gels that improve the separation of proteins and their detection by silver staining. *Analytical Biochemistry*, **173**, 412-423.
- HOCHSTRASSER, A.C., JAMES, R.W., POMETTA, D. AND HOCHSTRASSER, D.F. (1991a). Preparative isoelectrofocusing and high resolution two-dimensional electrophoresis for concentration and purification of proteins. *Applied and Theoretical Electrophoresis*, **1**, 333-337.
- HOCHSTRASSER, D.F., APPEL, R.D., VARGAS, R., PERRIER, R., VUROLLO, J.F., RAVIER, F., PASQUALI, C., FUNK, M., PELLIGRINI, C., MÜLLER, A.F. AND SCHIERRER, J.R. (1991b). A clinical molecular scanner: the Melanie project. *Medical Computing*, **8**, 85-91.
- HOCHSTRASSER, D.F., FRUTIGER, S., PAQUET, N., BAÏROCH, A., RAVIER, F., PASQUALI, C., SANCHEZ, J.-C., TISSOT, J.-D., BJELLQVIST, B., VARGAS, R., APPEL, R.D. AND HUGHES, G.J. (1992). Human liver protein map: a reference database established by microsequencing and gel comparison. *Electrophoresis*, **13**, 992-1001.
- HOLT, T.G., CHANG, C., LAURENT-WINTER, C., MURAKAMI, T., DAVIES, J.E. AND THOMPSON, C.J. (1992). Global changes in gene expression related to antibiotic synthesis in *Streptomyces hygroscopicus*. *Molecular Microbiology*, **6**, 969-980.
- HONORE, B., LEFFERS, H., MADSEN, P. AND CELIS, J.E. (1993). Interferon-gamma up-regulates a unique set of proteins in human keratinocytes. Molecular cloning and expression of the cDNA encoding the RGD-sequence containing protein IGUP 1-5111. *European Journal of Biochemistry*, **218**, 421-430.
- HUBERTY, M.C., VATH, J.E., YU, W. AND MARTIN, S.A. (1993). Site-specific carbohydrate

- identification in recombinant proteins using MALD-TOF MS. *Analytical Chemistry*, **65**, 2791-2800.
- HUGHES, G.J., FRUTIGER, S., PAQUET, N., PASQUALI, C., SANCHEZ, J.-C., TISSOT, J.D., BAIROCH, A., APPEL, R.D. AND HOCHSTRASSER, D.F. (1993). Human liver protein map: update 1993. *Electrophoresis*, **14**, 1216-1222.
- HUGHES, J.H., MACK, K. AND HAMPARIAN, V.V. (1988). India ink staining of proteins on nylon and hydrophobic membranes. *Analytical Biochemistry*, **173**, 18-25.
- JAMES, P., QUADRONI, M., CARAFOLI, E. AND GONNET, G. (1993). Protein identification by mass profile fingerprinting. *Biochemical and Biophysical Research Communications*, **195**, 58-64.
- JL, H., WHITEHEAD, R.H., REID, G.E., MORITZ, F.L., WARD, L.D. AND SIMPSON, R.J. (1994). Two-dimensional electrophoretic analysis of proteins expressed by normal and cancerous human crypts: application of mass spectrometry to peptide-mass fingerprinting. *Electrophoresis*, **15**, 391-405.
- JOHNSON, R.S. AND WALSH, K.A. (1992). Sequence analysis of peptide mixtures by automated integration of Edman and mass spectrometric data. *Protein Science*, **1**, 1083-1091.
- JOHNSTON, R.F., PICKETT, S.C. AND BARKER, D.L. (1990). Autoradiography using storage phosphor technology. *Electrophoresis*, **11**, 355-360.
- JUNGBLUT, P., DZIONARA, M., KLOSE, J. AND WITTMANN-LEIBOLD, B. (1992). Identification of tissue proteins by amino acid analysis after purification by two-dimensional electrophoresis. *Journal of Protein Chemistry*, **11**, 603-612.
- JUNGBLUT, P., OTTO, A., ZEINDL-EBERHART, E., PLEIßNER, K.-P., KNECHT, M., REGITZ-ZAGROSEK, V., FLECK, E. AND WITTMANN-LEIBOLD, B. (1994). Protein composition of the human heart: the construction of a myocardial two-dimensional electrophoresis database. *Electrophoresis*, **15**, 685-707.
- KOHARA, Y., AKIYAMA, K. AND ISONO, K. (1987). The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. *Cell*, **50**, 495-508.
- KLOSE, J. (1975). Protein mapping by combined isoelectric focusing and electrophoresis in mouse tissues. A novel approach to testing for individual point mutations in mammals. *Human Genetik*, **26**, 231-243.
- LATHAM, K.E., GARRELS, J.I., CHANG, C. AND SOLTER, D. (1991). Quantitative analysis of protein synthesis in mouse embryos I: extensive re-programming at the one- and two-cell stages. *Development*, **2**, 921-932.
- LATHAM, K.E., GARRELS, J.I., CHANG, C. AND SOLTER, D. (1992). Analysis of embryonic mouse development: construction of a high-resolution, two-dimensional gel protein database. *Applied and Theoretical Electrophoresis*, **2**, 163-170.
- LATHAM, K.E., GARRELS, J.I. AND SOLTER, D. (1993). Two-dimensional analysis of protein synthesis. *Methods in Enzymology*, **255**, 473-489.
- LE MAIRE, M., DESCHAMPS, S., MOLLER, J.V., LE CAER, J.P. AND ROSSIER, J. (1993). Electrospray ionization mass spectrometry from sodium dodecyl sulfate-polyacrylamide gel electrophoresis: application to the topology of the sarcoplasmic reticulum Ca^{2+} -ATPase. *Analytical Biochemistry*, **214**, 50-57.
- LEMKIN, P.F. AND LESTER, E.P. (1989). Database and search techniques for two-dimensional gel protein data: a comparison of paradigms for exploratory data analysis and prospects for biological modelling. *Electrophoresis*, **10**, 122-140.
- LEMKIN, P.F., WU, Y. AND UPTON, K. (1993). An efficient disk-based data structure for rapid searching of quantitative two-dimensional gel databases. *Electrophoresis*, **14**, 1341-1350.
- LI, K.W., GERAERTS, W.P., VAN-ELK, R. AND KOOSE, J. (1989). Quantification of proteins in the subnanogram and nanogram range: comparison of the AuroDye, FerriDye, and india ink staining methods. *Analytical Biochemistry*, **182**, 44-47.
- LIANG, P. AND PARDEE, A.B. (1992). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science*, **257**, 967-971.
- MANN, M. (1995). Sequence database searching by mass spectrometric data. In *Microcharacterisation of Proteins* (R. Kellner, F. Lottspeich, and H.E. Meyer, Eds.), pp 223-245. VCH, Weinheim.

- MANN, M., HOJRUP, P. AND ROEPSTORFF, P. (1993). Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biological Mass Spectrometry*, **22**, 338-345.
- MANN, M. AND WILM, M. (1994). Error tolerant identification of peptides in sequence databases by peptide sequence tags. *Analytical Chemistry*, **66**, 4390-4399.
- MATSUDAIRA, P. (1987). Sequence of picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *Journal of Biological Chemistry*, **262**, 10035-10038.
- MONARDO, P.J., BOUTELL, T., GARRELS, J.I. AND LATTER, G.I. (1994). A distributed system for two-dimensional gel analysis. *Computer Applications in the Biosciences*, **10**, 137-143.
- MORTZ, E., VORM, O., MANN, M. AND ROEPSTORFF, P. (1994). Identification of proteins in polyacrylamide gels by mass spectrometric peptide mapping combined with database search. *Biological Mass Spectrometry*, **23**, 249-261.
- MURTHY, L.R. AND IQBAL, K. (1991). Measurement of picomoles of phosphoamino acids by high performance liquid chromatography. *Analytical Biochemistry*, **193**, 299-305.
- MYRICK, J.E., LEMKIN, P.F., ROBINSON, M.K. AND UPTON, K.M. (1993). Comparison of the BiImage Visage 2000 and the GELLAB-II two-dimensional electrophoresis image analysis systems. *Applied and Theoretical Electrophoresis*, **3**, 335-346.
- NEIDHARDT, F.C., APPLEBY, D.B., SANKAR, P., HUTTON, M.E. AND PHILLIPS, T.A. (1989). Genomically linked cellular protein databases derived from two-dimensional polyacrylamide gel electrophoresis. *Electrophoresis*, **10**, 116-122.
- NIKODEM, V. AND FRESCO, J.R. (1979). Protein fingerprinting by SDS-gel electrophoresis after partial fragmentation with CNBr. *Analytical Biochemistry*, **97**, 382-386.
- NOKIHARA, K., MORITA, N. AND KURIKI, T. (1992). Applications of an automated apparatus for two-dimensional electrophoresis. Model TEP-1, for microsequence analysis of proteins. *Electrophoresis*, **13**, 701-707.
- O'FARRELL, P.H. (1975). High resolution two-dimensional electrophoresis of proteins. *Journal of Biological Chemistry*, **250**, 4007-4021.
- O'FARRELL, P.Z., GOODMAN, H.M. AND O'FARRELL, P.H. (1977). High resolution two-dimensional electrophoresis of basic as well as acidic proteins. *Cell*, **12**, 1133-1142.
- OLIVER *et al.* (1992). The complete DNA sequence of yeast chromosome III. *Nature* **357**, 38-46.
- OLSEN, A.D. AND MILLER, M.J. (1988). Elsie 4: quantitative computer analysis of sets of two-dimensional gel electrophoretograms. *Analytical Biochemistry*, **169**, 49-70.
- ORTIZ, M.L., CALERO, M., FERNANDEZ-PATRON, C., PATRON, C.F., CASTELLANOS, L. AND MENDEZ, E. (1992). Imidazole-SDS-Zn reverse staining of proteins in gels containing or not SDS and microsequence of individual unmodified electroblotted proteins. *FEBS Letters*, **296**, 300-304.
- OSTERGREN, K., ERIKSSON, G. AND BJELLQVIST, B. (1988). The influence of support material used on band sharpness in immobiline gels. *Journal of Biochemical and Biophysical Methods*, **16**, 165-170.
- OL, K., WILKINS, M.R., YAN, J.X., GOOLEY, A.A., FUNG, Y., SHEUMACK, D. AND WILLIAMS, K.L. (1995). Improved high-performance liquid chromatography of amino acids derivatised with 9-fluorenylmethyl chloroformate. *Journal of Chromatography* (in press).
- PACKER, N., WILKINS, M.R., GOLAZ, O., LAWSON, M., GOOLEY, A.A., HOCHSTRASSER, D.F., REDMOND, J. AND WILLIAMS, K.L. (1995). Characterisation of human plasma glycoproteins separated by two-dimensional gel electrophoresis. *BioTechnology* (in press).
- PAPPIN, D.J.C., HOJRUP, P. AND BLEASBY, A.J. (1993). Rapid identification of proteins by peptide-mass fingerprinting. *Current Biology*, **3**, 327-332.
- PATTERSON, S.D. (1994). From electrophoretically separated protein to identification: strategies for sequence and mass analysis. *Analytical Biochemistry*, **221**, 1-15.
- PATTERSON, S.D. AND LATTER, G.I. (1993). Evaluation of storage phosphor imaging for quantitative analysis of 2-D gels using the Quest II system. *BioTechniques*, **15**, 1076-1083.
- PISANO, A., REDMOND, J.W., WILLIAMS, K.L. AND GOOLEY, A.A. (1993). Glycosylation sites identified by solid-phase Edman degradation: O-linked glycosylation motifs on human glycoporphin A. *Glycobiology*, **3**, 429-435.
- RABILLOU, T. (1992). A comparison between low background silver diamine and silver nitrate protein stains. *Electrophoresis*, **13**, 429-439.

- RASMUSSEN, H.H., VAN DAMME, J., PUYPE, M., GESSER, B., CELIS, J.E. AND VANDEKERCKHOVE, J. (1992). Microsequences of 145 proteins recorded in the two-dimensional gel protein database of normal human epidermal keratinocytes. *Electrophoresis*, **13**, 960-969.
- RASMUSSEN, H.H., MORTZ, E., MANN, M., ROEPSTORFF, P. AND CELIS, J.E. (1994). Identification of transformation sensitive proteins recorded in human two-dimensional gel protein databases by mass-spectrometric peptide mapping alone and in combination with microsequencing. *Electrophoresis*, **15**, 406-416.
- RICHARDSON, F.C., HORN, D.M. AND ANDERSON, N.L. (1994). Dose-responses in rat hepatic protein modification and expression following exposure to the rat hepatocarcinogen methapyrilene. *Carcinogenesis*, **15**, 325-329.
- RIGHETTI, P.G. (1990). Immobilized pH gradients: theory and methodology. In *Laboratory Techniques in Biochemistry and Molecular Biology* (R.H. Burdon and P.H. van Knippenberg, Eds) Elsevier, Amsterdam.
- RIGHETTI, P.G. AND DRYSDALE, J.W. (1973). *Annals of the New York Academy of Sciences*, **209**, 163-186.
- RODRIGUEZ, L.V., GERNSTEN, D.M., RAMAGLI, L.S. AND JOHNSTON, D.A. (1993). Towards stoichiometric silver staining of proteins resolved in complex two-dimensional electrophoresis gels: real-time analysis of pattern development. *Electrophoresis*, **14**, 628-637.
- ROSENFELD, J., CAPDEVIELLE, J., GUILLEMOT, J.C. AND FERRARA, P. (1992). In-gel digestion of proteins for internal sequence analysis after one- or two-dimensional gel electrophoresis. *Analytical Biochemistry*, **203**, 173-179.
- SANCHEZ, J.C., RAVIER, F., PASQUALI, C., FRUTIGER, S., PAQUET, N., BJELLOVIST, B., HOCHSTRASSER, D.F. AND HUGHES, G.J. (1992). Improving the detection of proteins after transfer to polyvinylidene difluoride membranes. *Electrophoresis*, **13**, 715-717.
- SANGER, F., COULSON, A.R., HONG, G.F., HILL, D.F. AND PETERSEN, G.B. (1982). Nucleotide sequence of bacteriophage λ DNA. *Journal of Molecular Biology*, **162**, 729-773.
- SCHEELE, G.J. (1975). Two-dimensional analysis of soluble proteins. *Biochemistry*, **250**, 5375-5385.
- SHAW, G. (1993). Rapid identification of proteins. *Proceedings of the National Academy of Sciences USA*, **90**, 5138-5142.
- SIBBALD, P.R., SOMMERFELDT, H. AND ARGOS, P. (1991). Identification of proteins in sequence databases from amino acid composition. *Analytical Biochemistry*, **198**, 330-333.
- SIMPSON, R.J., TSUGITA, A., CELIS, J.E., GARRELS, J.I. AND MEWES, H.W. (1992). Workshop on two-dimensional gel protein databases. *Electrophoresis*, **13**, 1055-1061.
- SINHA, P.K., KOTTGEN, E., STOFFLER, M.-M., GIANAZZA, E. AND RIGHETTI, P.G. (1990). Two-dimensional maps in very acidic immobilized pH gradients. *Journal of Biochemical and Biophysical Methods*, **20**, 345-352.
- SMITH, D.W. (1994). Introduction. In *Biocomputing: Informatics and Genome Projects* (D.W. Smith, Ed.), ppl-12. Academic Press, San Diego.
- STRUPAT, K., KARAS, M., HILLENKAMP, F., ECKERSKORN, C. AND LOTTSPREICH, F. (1994). Matrix-assisted laser desorption/ionization mass spectrometry of proteins electroblotted after polyacrylamide gel electrophoresis. *Analytical Chemistry*, **66**, 464-470.
- SUTTON, C.W., PEMBERTON, K.S., COTTRELL, J.S., CORBETT, J.M., WHEELER, C.H., DUNN, M.J. AND PAPPIN, D.J. (1995). Identification of myocardial proteins from two-dimensional gels by peptide mass fingerprinting. *Electrophoresis*, **16**, 308-316.
- TOUS, G.I., FAUSNAUGH, J.L., AKINYOSOYE, O., LACKLAND, H., WINTERCASH, P., VITORICA, F.J. AND STEIN, S. (1989). Amino acid analysis on polyvinylidene difluoride membranes. *Analytical Biochemistry*, **179**, 50-55.
- TOVEY, E.R., FORD, S.A. AND BALDO, B.A. (1987). Protein blotting on nitrocellulose: some important aspects of the resolution and detection of antigens in complex extracts. *Journal of Biochemical and Biophysical Methods*, **14**, 1-17.
- URWIN, V.E. AND JACKSON, P. (1993). Two-dimensional polyacrylamide gel electrophoresis of proteins labeled with the fluorophore monobromobimane prior to first-dimensional isoelectric focusing: imaging of the fluorescent protein spot patterns using a cooled charge-coupled device. *Analytical Biochemistry*, **209**, 57-62.
- VANBOGELEN, R.A., HUTTON, M.E. AND NEIDHARDT, F.C. (1990). Gene-protein database

- of *Escherichia coli* K-12: edition 3. *Electrophoresis*, **11**, 1131–1166.
- VANBOGELEN, R.A. AND NEIDHARDT, F.C. (1991). The gene-protein database of *Escherichia coli*: edition 4. *Electrophoresis*, **12**, 955–994.
- VANBOGELEN, R.A., SANKER, P., CLARK, R.L., BOGAN, J.A. AND NEIDHARDT, F.C. (1992). The gene-protein database of *Escherichia coli*: edition 5. *Electrophoresis*, **13**, 1014–1054.
- VANDEKERKHOVE, J., BAUW, G., VANCOMPERNOLLE, K., HONORE, B. AND CELIS, J. (1991). Comparative two-dimensional gel analysis and microsequencing identifies gelsolin as one of the most prominent downregulated markers of transformed human fibroblast and epithelial cells. *Journal of Cell Biology*, **111**, 95–102.
- VANFLETEREN, J.R., RAYMACKERS, J.G., VAN BUN, S.M. AND MEHUS, L.A. (1992). Peptide mapping and microsequencing of proteins separated by SDS-PAGE after limited *in situ* hydrolysis. *BioTechniques*, **12**, 550–557.
- VORM, O. AND MANN, M. (1994). Improved mass accuracy in matrix-assisted laser desorption/ionization time-of-flight mass spectrometry of peptides. *Journal of the American Society for Mass Spectrometry*, **5**, 955–958.
- VORM, O., ROEPSTORFF, P. AND MANN, M. (1994). Improved resolution and very high sensitivity in MALDI TOF of matrix surfaces made by fast evaporation. *Analytical Chemistry*, **66**, 3281–3287.
- WALLACE, A. AND SALLZ, H.P. (1992a). Ultramicrodetection of proteins in polyacrylamide gels. *Analytical Biochemistry*, **203**, 27–34.
- WALLACE, A. AND SALLZ, H.P. (1992b). Beyond silver staining. *Nature*, **357**, 608–609.
- WALSH, B.J., GOOLEY, A.A., WILLIAMS, K.L. AND BREIT, S.N. (1995). Identification of macrophage activation associated proteins by two-dimensional electrophoresis and microsequencing. *Journal of Leukocyte Biology*, **57**, 507–512.
- WASINGER, V.C., CORDWELL, S.J., POLJAK, A., YAN, J.X., GOOLEY, A.A., WILKINS, M.R., DUNCAN, M., HARRIS, R., WILLIAMS, K.L. AND HUMPHERY-SMITH, I. (1995). Progress with Gene-Product Mapping of the Mollicutes: *Mycoplasma genitalium*. *Electrophoresis*, **16**. In Press.
- WEITZHANDLER, M., KADLECEK, D., AVDALOVIC, N., FORTE, J. G., CHOW, D. AND TOWNSEND, R. R. (1993). Monosaccharide and oligosaccharide analysis of proteins transferred to polyvinylidene fluoride membranes after sodium dodecyl sulfate-polyacrylamide gel electrophoresis. *Journal of Biological Chemistry*, **268**, 5121–5130.
- WILKINS, M.R., PASQUALI, C., APPEL, R.D., OU, K., GOLAZ, O., SANCHEZ, J-C., YAN, J.X., GOOLEY, A.A., HUGHES, G., HUMPHERY-SMITH, I., WILLIAMS, K.L. AND HOCHSTRASSER, D.F. (1995). From Proteins to Proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. Submitted.
- WILKINS, M.R., OU, K., APPEL, R.D., GOLAZ, O., PASQUALI, C., YAN, J.X., FARNSWORTH, V., CARTIER, P., HOCHSTRASSER, D.F., WILLIAMS, K.L. AND GOOLEY, A.A. (1996). Rapid protein identification using N-terminal sequence tagging and amino acid analysis (submitted).
- WIRTH, P.J., LUO, L.D., FUJIMOTO, Y., BISGAARD, H.C. AND OLSEN A.D. (1991). The rat liver epithelial (RLE) cell protein database. *Electrophoresis*, **12**, 931–954.
- WIRTH, P.J., LUO, L.D., BENJAMIN, T., HUANG, T.N., OLSEN A.D. AND PARMALEE, D.C. (1993). The rat liver epithelial (RLE) cell nuclear protein database. *Electrophoresis*, **14**, 1199–1215.
- WU, Y., LEMKIN, P.F. AND UPTON, K. (1993). A fast spot segmentation algorithm for two-dimensional gel electrophoresis analysis. *Electrophoresis*, **14**, 1351–1356.
- YAMAGUCHI, K. AND ASAKAWA, H. (1988). Preparation of colloidal gold for staining proteins electrotransferred onto nitrocellulose membranes. *Analytical Biochemistry*, **172**, 104–107.
- YAMASHITA, K., IDEO, H., OHKURA, T., FUKUSHIMA, K., YUASA, I., OHNO, K. AND TAKESHITA, K. (1993). Sugar chains of serum transferrin from patients with carbohydrate deficient glycoprotein syndrome. Evidence of asparagine-N-linked oligosaccharide transfer deficiency. *Journal of Biological Chemistry*, **268**, 5783–5789.
- YATES, J.R. III, SPEICHER, S., GRIFFIN, P.R. AND HUNKAPILLER, T. (1993). Peptide mass maps: a highly informative approach to protein identification. *Analytical Biochemistry*, **214**, 397–408.

3 The

R.M. D.

Therm.
Hamilton
Wales 2
Private

Introdu

Proteas
dealing
type. T
structur
since th
by Was
1930) c
possibly
mechan
subtilis
1986; f
comme
proteas
the cha
been in
regulat
apparai
Kalisz.
Barrett
Give
surpris
attracti
subject
stabilit
enzyme
Sherod

* Corres

Biotechn
0264-872

Human cellular protein patterns and their link to genome DNA sequence data: usefulness of two-dimensional gel electrophoresis and microsequencing

JULIO E. CELIS,*¹ HANNE H. RASMUSSEN,* HENRIK LEFFERS,* PEDER MADSEN,* BENT HONORÉ,* BORBALA GESSER,* KURT DEJGAARD,* JOËL VANDEKERCKHOVE*

¹*Institute of Medical Biochemistry and Human Genome Research Centre, Aarhus University, DK-8000 Aarhus C, Denmark, and* ^{*}*Laboratorium voor Fysiologische Chemie, Rijksuniversiteit Gent, Belgium*

ABSTRACT Analysis of cellular protein patterns by computer-aided 2-dimensional gel electrophoresis together with recent advances in protein sequence analysis have made possible the establishment of comprehensive 2-dimensional gel protein databases that may link protein and DNA information and that offer a global approach to the study of the cell. Using the integrated approach offered by 2-dimensional gel protein databases it is now possible to reveal phenotype specific protein (or proteins), to microsequence them, to search for homology with previously identified proteins, to clone the cDNAs, to assign partial protein sequence to genes for which the full DNA sequence and the chromosome location is known, and to study the regulatory properties and function of groups of proteins that are coordinately expressed in a given biological process. Human 2-dimensional gel protein databases are becoming increasingly important in view of the concerted effort to map and sequence the entire genome. — Celis, J. E.; Rasmussen, H. H.; Leffers, H.; Madsen, P.; Honoré, B.; Gesser, B.; Dejgaard, K.; Vandekerckhove, J. Human cellular protein patterns and their link to genome DNA sequence data: usefulness of two-dimensional gel electrophoresis and microsequencing. *FASEB J.* 5: 2200-2208; 1991.

Key Words: human protein patterns • 2-dimensional gel protein databases • gene expression • microsequencing • cDNA cloning • linking protein and DNA information • genome mapping and sequencing

PROTEINS SYNTHESIZED FROM information contained in the DNA orchestrate most cellular functions. The total number of proteins synthesized by a typical human cell is unknown although current estimates range from 3000 to 6000. Of these, as many as 70% may perform household functions and are expected to be shared by all cell types irrespective of their origin. There are many different cell types in the human body with perhaps 30,000 to 50,000 proteins expressed in the organism as a whole judged from the fact that about 3% of the haploid genome correspond to genes. Today only a small fraction of the total set of proteins has been identified, and little is known about the protein patterns of individual cell types or their variation under physiological and abnormal conditions.

For the past 15 years, high resolution 2-dimensional gel electrophoresis has been the technique of choice to determine the protein composition of a given cell type and for monitoring changes in gene activity through quantitative and qualitative analysis of the thousands of proteins that orchestrate various cellular functions (refs 1-6 and references

therein). The technique originally described by O'Farrell separates proteins in terms of their isoelectric point (pI) and molecular weight. Usually one chooses a condition of interest and the cell reveals the global protein behavioral response as all detected proteins can be analyzed both qualitatively and quantitatively in relation to each other. At present, most available 2-dimensional gel techniques (regular gel format) can resolve between 1000 and 2000 proteins from a given mammalian cell type, a number that corresponds to about 2 million base pairs of coded DNA. Less abundant proteins can be detected by analyzing partially purified cellular fractions.

Two-dimensional gel electrophoresis has been widely applied to analysis of cellular protein patterns from bacteria to mammalian cells (refs 1-6, and references therein). In spite of much work, however, information gathered from these studies has not reached the scientific community in its fullness because of lack of standardized gel systems and the lack of means for storing and communicating protein information. Only recently, because of the development of appropriate computer software (7-13), has it been possible to scan gels, assign numbers to individual proteins, and store the wealth of information in quantitative and qualitative comprehensive 2-dimensional gel protein databases (4, 14-23), i.e., those containing information about the various properties (physical, chemical, biological, biochemical, physiological, genetic, immunological, architectural, etc.) of all the proteins that can be detected in a given cell type. Such integrated 2-dimensional gel protein databases offer an easy and standardized medium in which to store and communicate protein information and provide a unique framework in which to focus a multidisciplinary approach to study the cell. Once a protein is identified in the database, all of the information accumulated can be easily retrieved and made available to the researcher. In the long run, protein databases are expected to foster a wide variety of biological information that may be instrumental to researchers working in many areas of biology—among others, cancer and oncogene studies, differentiation, development, drug development and testing, genetic variation, and diagnosis of genetic and clinical diseases (Fig. 1).

The approach using systematic 2-dimensional gel protein analysis has recently gained a new dimension with the advent of techniques to microsequence major proteins recorded

¹To whom correspondence should be addressed, at: Institute of Medical Biochemistry and Human Genome Research Centre, Ole Worms Alle, Bldg. 170, University Park, DK-8000 Aarhus C, Denmark.

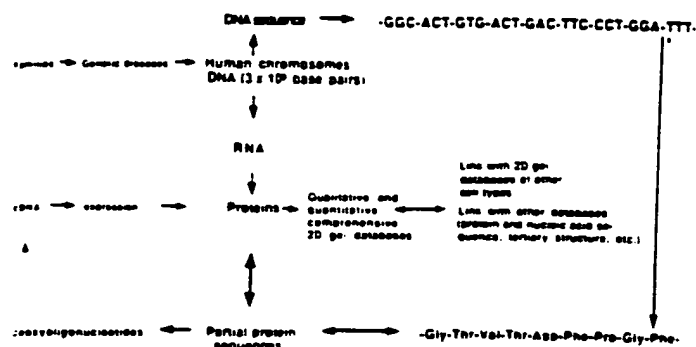


Figure 1. Interface between partial protein sequence databases, comprehensive 2-dimensional gel databases, and the human genome sequencing project. Appropriate software is required to compare protein and DNA sequences. In general, although the inference of a protein's sequence from the DNA sequence (thick arrow) is direct and unambiguous, the DNA sequence can only be inferred approximately from the protein sequence (thin arrow) and cloning of the gene requires either a cDNA or the requisite group of oligonucleotide probes deduced from the partial amino acid sequence. Modified from ref. 6.

in the databases (refs 24-42 and references therein). Partial protein sequences can be used to search for protein identity as well as to prepare specific DNA probes for cloning as yet-uncharacterized proteins (Fig. 1). As these sequences can be stored in the database (see for example Fig. 2H), they offer a unique opportunity to link information on proteins with the existing or forthcoming DNA sequence data on the human genome (Fig. 1) (20, 36, 39).

Using the integrated approach offered by comprehensive 2-dimensional gel databases (Fig. 1), it will be possible to identify phenotype-specific proteins; microsequence them and store the information in the database; search for homology with previously characterized proteins; clone the cDNAs, assign partial protein sequences to genes for which the full DNA sequence and the chromosome location are known, and study the regulatory properties and function of groups of proteins (pathways, organelles, etc.) that are coordinately expressed in a given biological process. Comprehensive 2-dimensional gel protein databases will depict an integrated picture of the expression levels and properties of the thousands of protein components of organelles, pathways, and cytoskeletal systems in both physiological and abnormal conditions and are expected to lead to identification of new regulatory networks in different cell types and organisms. In the future, 2-dimensional gel protein databases may be linked to each other as well as to national and international specialized databanks on nucleic acid and protein sequences, protein structures, NMR experimental data, complex carbohydrates, etc.

A few 2-dimensional gel protein databases that are accessible in a computer form have been published in extenso: these correspond to the protein-gene database of *Escherichia coli* K-12 developed by Neidhardt and colleagues (14, 23), the rat REF 52 database established by Garrels and co-workers at Cold Spring Harbor (18, 22), and a few human databases (transformed amnion cells [15, 20], normal embryonal lung MRC-5 fibroblasts [17, 21], keratinocytes [19] and peripheral blood mononuclear cells [15]) developed in Aarhus. Given space limitations and to keep this review in focus, we will concentrate on the computerized analysis of human cellular 2-dimensional gel patterns, and in particular on the steps involved in establishing comprehensive 2-dimensional gel databases that can link protein and DNA information.

MAKING AND MANAGING A COMPREHENSIVE 2-DIMENSIONAL GEL DATABASE OF HUMAN CELLULAR PROTEINS

The first step in making a comprehensive 2-dimensional gel protein database is to prepare a synthetic image (digital form of the gel image) of the gel (fluorogram, Coomassie blue or silver stained gel) to be used as a standard or master reference. This can be done with laser scanners, charge couple device (CCD)² array scanners, television cameras, rotating drum scanners, and multiwire chambers (15). Computerized analysis systems for spot detection, quantitation, pattern matching, and data handling (access and retrieval of information, database making) have been described in the literature (ELSIE [43], GELLAB [11], HERMES [44], MELANIE [10], QUEST (9), and TYCHO [8]) and some are available commercially (PDQUEST, Protein Database Inc., Huntington, N.Y.; KEPLER, Large Scale Biology, Rockville, Md.; Visage, BioImage Corporation, Ann Arbor, Mich.; Gemini, Joyce Loebl, Gateshead; Microscan 1000, Technology Resources Inc., Nashville, Tenn. and MasterScan, Billerica, Mass.). Unfortunately, most of these systems are incompatible with one another and their advantages and disadvantages have been discussed by Miller (13).

In our work station in Aarhus, fluorograms are scanned with a Molecular Dynamics laser scanner and the data are analyzed using the PDQUEST II software (Protein Databases Inc.) (12) running on a spark station computer 4100 FC-8-P3 from SUN Microsystems, Inc. The scanner measures intensity in the range of 0-2.0 absorbance. A typical scan of a 17 x 17 cm fluorogram takes about 2 min. Steps in image analysis include: initial smoothing, background subtraction, final smoothing, spot detection, and fitting of ideal Gaussian distribution to spot centers. Spot intensity is calculated as the integration of a fitted Gaussian. If calibration strips containing individual segments of a known amount of radioactivity are used, it is possible to merge multiple exposures of the sample image into a single data image of greater dynamic range. Once the synthetic image is created it can be stored on disk and displayed directly on the monitor. Functions that can be used to edit the images include: cancel (for example, to erase scratches that may have been interpreted as spots by the computer; cancel streaks or low dpm spots), combine (sometimes a spot may be resolved into several closely packed spots), restore, uncombine, and add spot to the gel. The process is time consuming—about 1-1/2 day per image. Edited standard images can be matched to other synthetic images. Figure 2A shows a portion of a standard synthetic image (IEF) of a fluorogram of [³⁵S]methionine labeled cellular proteins from human AMA cells (master database) (20). Images can be displayed either in black and white (resembling the original fluorograms) or in color (other images in Fig. 2), depending on the need. As shown in Fig. 2B, each polypeptide is assigned a number by the computer, which facilitates the entry and retrieval of qualitative and quantitative information for any given spot in the gel (20). The standard image can be matched automatically by the computer to other standard or reference gels (Fig. 2C, matching of AMA cellular proteins [left] to MRC-5 proteins [right]) provided a few landmark spots are given manually as reference (indicated with a + in Fig. 2C) to initiate the process.

²Abbreviations: CCD, charge couple device; PCNA, proliferating cell nuclear antigen; HPLC, high performance liquid chromatography.

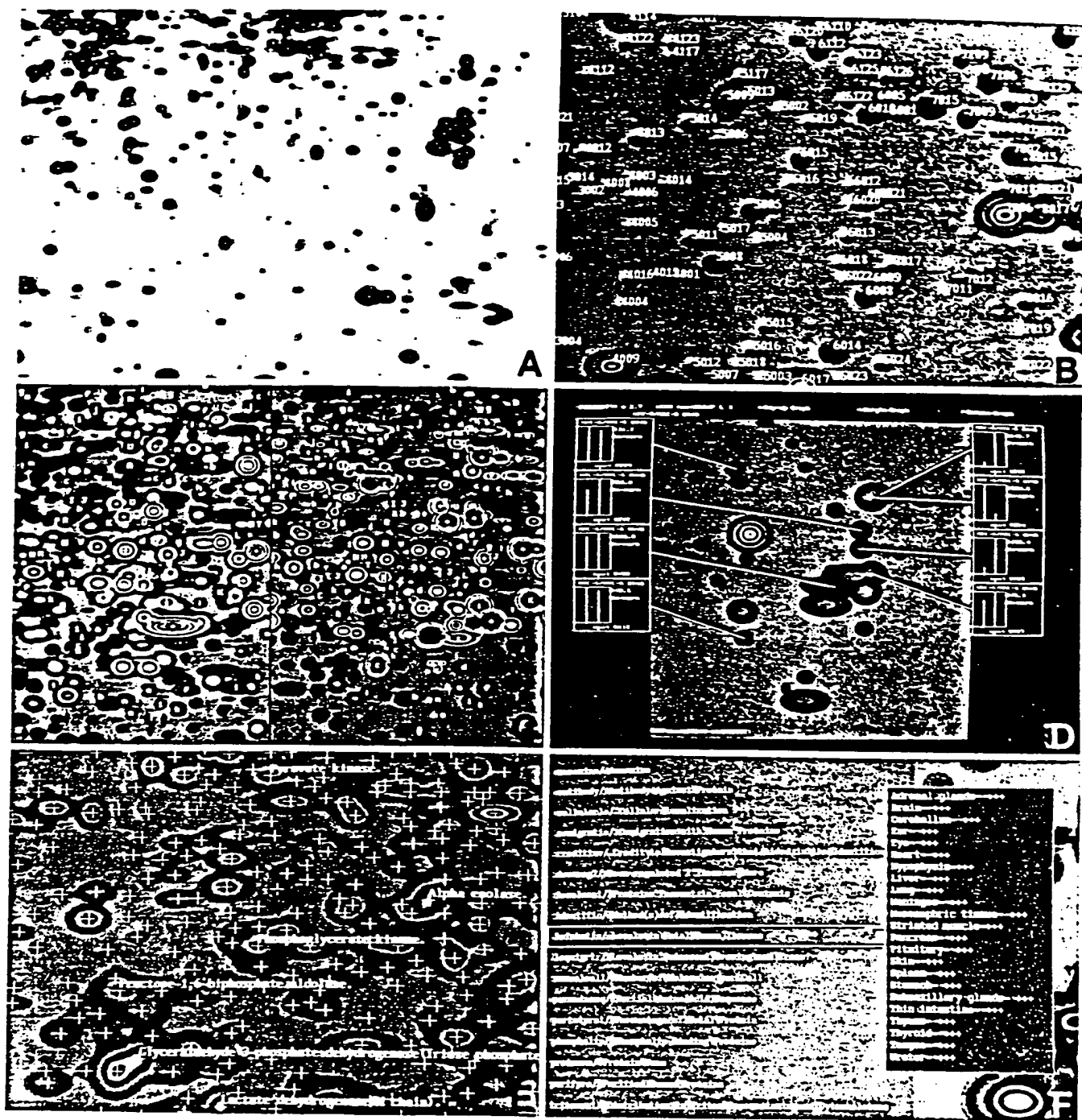
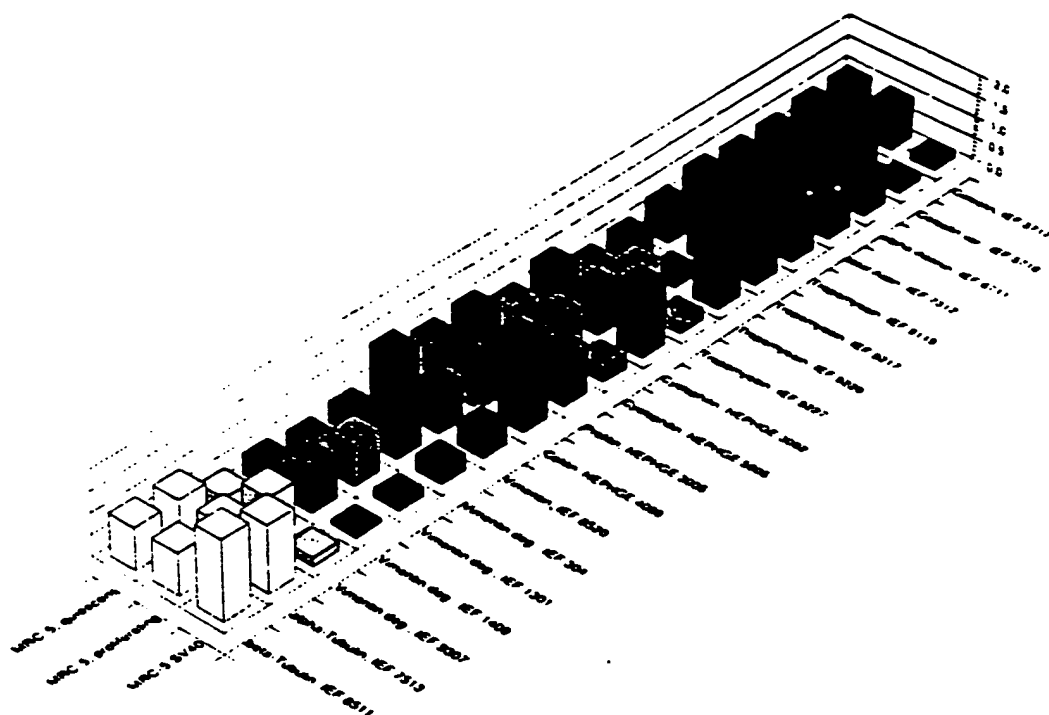
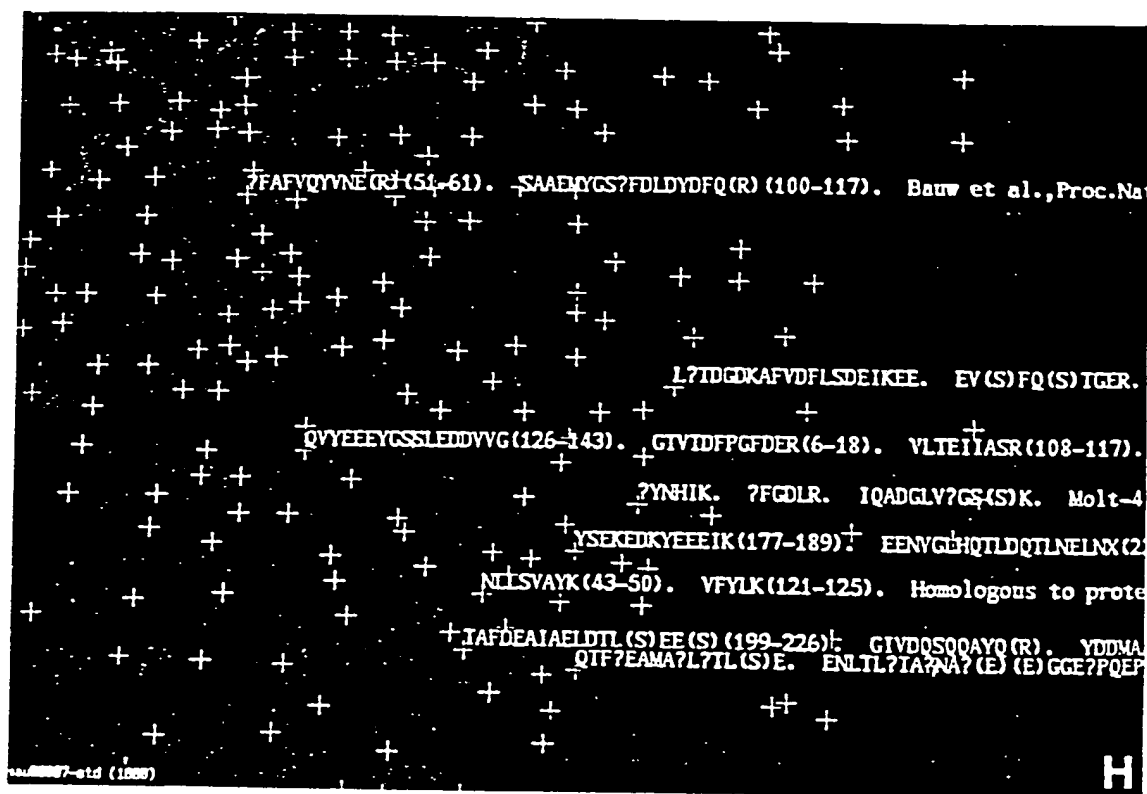


Figure 2. *A*: Synthetic image of a fraction of an IEF gel of the master image of AMA cellular proteins. *B*: As in *A* but showing numbers assigned to each spot. *C*: Comparison of AMA (left) and normal human embryonal lung MRC-5 fibroblasts (right) IEF proteins patterns. Matched proteins are indicated by a - or by the same letters in both gels. Once a protein is matched, information contained in the various categories available in the master AMA database can be transferred. *D*: Synthetic image of a fraction of an IEF fluorogram of [³⁵S]methionine-labeled proteins from normal human MRC-5 fibroblasts. The histograms show levels of synthesis of a few proteins in MRC-5 (left bar) and SV40 transformed MRC-5 (right bar) fibroblasts. *E*: Polypeptides that contain information under the category glycolytic pathway. *F*: The function peruse annotation for spot allows the operator to inquire about categories and information available for a given protein. *G*: Relative abundance of cytoskeletal and cytoskeletal-related proteins in quiescent, proliferating, and SV40-transformed MRC-5 fibroblasts. *H*: Polypeptides that contain information under the category partial amino acid sequences.



G



H

The automatic matching process that has been described in detail by Garrels et al. (12) takes about 5 min. Matched proteins are indicated with the same letters in both gels (Fig. 2C). The usefulness of this function is emphasized by the fact that data accumulated on common household proteins can be easily transferred to any other human cellular cell type whose 2-dimensional gel cellular protein pattern is matched

to our standard AMA 2-dimensional gel protein image. Alternatively, if the standard gel is part of a matchset (set of gels in a given experiment) it can be used as a linker gel to compare, for example, the quantitative values of a given protein throughout the experiment (see Fig. 2D; levels of some proteins in normal and SV40 transformed human MRC-5 fibroblasts) or with other standard images in different sets of

cross-matched experiments (18, 22).

Once a standard map of a given protein sample is made, one can enter qualitative annotations to make a reference database. Our master 2-dimensional gel database of transformed human amnion cell (AMA) proteins (20) lists 3430 polypeptides of which 2592 correspond to cellular components, having pI's ranging from 4 to 13 and molecular weights between 8.5 and 230 kDa. The most abundant proteins in the database correspond to total actin (3.87% of total protein; about 90 million molecules per cell) while the lesser abundant of the recorded polypeptides are present in the vicinity of 5000 molecules per cell. Some annotation categories we are using to establish the master AMA database include: 1) protein identification (comigration with purified proteins, 2-dimensional immunoblotting, microsequencing); 2) amounts (total amounts and levels of synthesis); 3) subcellular localization (nuclear, cytoskeletal, membrane, membrane receptors, specific organelles, etc.); 4) antibodies; 5) posttranslational modifications (phosphorylation, glycosylation, methylation etc.); 6) microsequencing; 7) cell cycle specificity (specific variations in levels of synthesis and amount); 8) regulatory behavior (effect of hormones, growth factors, heat shock, etc.); 9) rate of synthesis in normal and transformed cells (proliferation sensitive proteins, cell cycle specific proteins, oncogenes, components of the pathway (or pathways) that control cell proliferation); 10) function (mainly from comigration with proteins of known function); 11) sets of proteins that are coordinately regulated (hierarchy of controls, differential gene expression in various cells, etc.); 12) cDNAs (cloned cDNAs); 13) proteins that are specific to a given disease (systematic comparison of protein patterns of fibroblast proteins from healthy and diseased individuals); 14) expression and exploitation of transfected cDNAs; 15) pathways (metabolic, others); 16) gene localization (genetic and physical); 17) effect of microinjected antibody on patterns of protein synthesis; and 18) secreted proteins.

Information entered for any spot in a given annotation category can be easily retrieved by asking the computer to display the information on the color screen. For example, Fig. 2E shows a synthetic image of a NEPHGE gel (master AMA database) displaying the information contained under the entry glycolytic pathway. Alternatively, one can use the function peruse annotations for spot to directly ask the computer to list all the entries available for a particular protein. By clicking the mouse in a given entry (in this case, presence in fetal human tissues) it is possible to take a quick look at the information in that particular entry (Fig. 2F).

A major obstacle encountered in building comprehensive 2-dimensional gel protein databases is identifying the large number of proteins separated by this technology. In our databases (20, 21), known proteins are identified by one or a combination of the following procedures: 1) comigration with known proteins, 2) 2-dimensional gel immunoblotting using specific antibodies, and 3) microsequencing of Coomassie Brilliant Blue stained human proteins recovered from dried 2-dimensional gels (see next section). Protein identification by means of microsequencing may be difficult, as individual protein members of families with short peptide differences may escape detection. In the gene-protein database of *E. coli* K-12 (14, 23), another major 2-dimensional gel database available at present, proteins are being identified by a wider range of tests that include comigration with purified proteins; genetic criterion (deletion, insertion, frameshift, nonsense, missense, regulatory), plasmid-bearing strains and in vitro synthesis of protein; selective labeling (methylation, phosphorylation); peptide map similarity; and physiological criterion and selective derivatization.

So far we have received nearly 550 antibodies from laboratories all over the world and these are being systematically tested by 2-dimensional gel immunoblotting for antigen determination. Similarly, purified proteins and organelles provided by several laboratories have greatly aided identification of unknown proteins (20, 21). We routinely request antibodies and protein samples and promise the donors to make available all the information we may have accumulated on that particular protein. For example, Table 1 lists entries available for Lipocortin V (IEF SSP 8216), also known as annexin V, VAC- α endonexin II, renocortin, chromobindin-5', anticoagulant protein, PAP-I, γ -calcimedlin, IBC, calphobindin, and anchorin CII.

As mentioned previously, one distinct advantage of 2-dimensional gel electrophoresis is the possibility of studying quantitative variations in cellular protein patterns that may lead to identification of groups of proteins that are expressed coordinately during a given biological process. Quantitation, however, is not an easy task as reflected by the lack of published data on global cellular protein patterns. We believe this is partly due to difficulties in obtaining sets of gels that are suitable for computer analysis (streaking, material remaining at the origin, etc.) as well as to limitations (laborious editing time, need of calibration strips to merge images, limited dynamic range, etc.) in the computer analysis systems available at the moment. Perhaps the most advanced quantitative studies published so far using computer analysis have been carried out by Garrels and co-workers (18, 22). In particular, these investigators have established a quantitative rat protein database (18, 22) designed to study growth control (proliferation, growth inhibitors, and stimulation) and transformation in well-defined groups of cell lines obtained by transformation of rat REF52 cells with SV40, adenovirus, and the Kirsten murine sarcoma virus. These studies have revealed clusters of proteins induced or repressed during growth to confluence as well as groups of transformation-sensitive proteins that respond in a differential fashion to transformation by DNA and RNA viruses. A most interesting feature of this quantitative database is the discovery of a group of coregulated proteins that show similar expression patterns as the cell cycle-regulated DNA replication protein known as proliferating cell nuclear antigen (PCNA)/cyclin (45).

In our human databases, most quantitations have been carried out by estimating the radioactivity contained in the polypeptides by direct counting of the gel pieces in a scintillation counter (20, 21). Up to 700 proteins can be cut out through appropriate exposed films in a period of time comparable to that required for editing a synthetic image. Manual quantitation of this large number of spots is difficult without the assistance of a master reference image and a numbering system that can be used to identify the spots. Using this approach, we have recorded quantitative changes in the relative abundance of 592 [³⁵S]methionine-labeled proteins synthesized by quiescent, proliferating, and SV40 transformed human embryonic lung MRC-5 fibroblasts (21). Some data concerning cytoskeletal and cytoskeletal-related proteins are presented in Fig. 2G. Our studies as well as those of Garrels and co-workers (18, 22) may in the long run help define patterns of gene expression that are characteristic of the transformed state.

OTHER 2-DIMENSIONAL GEL PROTEIN DATABASES

As mentioned previously there are other 2-dimensional gel databases available in computer form that have been pub-

TABLE 1. Some entries for lipocortin V in the human A.M.A. 2-dimensional gel protein database

Entries for lipocortin V (IEF SSP 8216)	Information entered
1. Protein name	Lipocortin V, renocortin, chromobindin-5', endonexin I, anticoagulant protein PAP-I, VAC- α , 35- γ -calcimedin, IBC, calphobindin I, anchorin CII, annexin V
2. Percentage of total protein	0.110% (about 2,800,000 molecules per cell)
3. Apparent molecular weight (mr)	33.3 kDa
4. Isoelectric point (pI)	4.76
5. Method (or methods) of identification	Microsequencing, 2-dimensional immunoblotting, Comigration
6. Credit to investigators that aided in identification	G. Bauw, J. Vandekerckhove, and colleagues, Rijksuniversiteit Gent; B. Pepinsky, BIOGEN, Cambridge; N.G. Ahn, University of Washington
7. Antibody against protein	Polyclonal (rabbit, antibody no. 20), B. Pepinsky, BIOGEN, Cambridge
8. Comigration with human proteins	Lipocortin V, N.G. Ahn, Howard Hughes Medical Institute, Washington University
9. Cellular localization	Subcortical membrane
10. Calcium-phospholipid-dependent membrane proteins	Lipocortin V
11. Function	Regulation of various aspects of inflammation, immune response, blood coagulation and differentiation
12. Partial amino acid sequence	GTVTDFPGFDER (7-18), VLTEIIASR (109-117), QVYEEENGSSLEDIDVVG (127-143), ?GTDEEKFITIFGT(R) (187-201)
13. cDNA sequence	Known, R. Blake et al., <i>J. Biol. Chem.</i> 263, 10799-10811; 1988 (pI = 4.76 from translated sequence)
14. Levels in fetal human tissues	Adrenal glands = + + +; brain = - - -; cerebellum = + + -; ear = - - -; eye = - - -; heart = - - -; hypophysis = + + +; liver = + + -; lung = + + +; meninges = + - -; mesonephric tissue = + + +; striated muscle = + + +; pancreas = - - -; skin = + - -; spleen = + + +; stomach = + - -; submandibular gland = + + +; small intestine = - - -; thymus = - - -; thyroid gland = + + +; tongue = - - -; ureter = - - -
15. Levels in quiescent, proliferating, and transformed MRC-5 fibroblasts	Q (quiescent) = 1.1; P (proliferating) = 1.0; T (SV40 transformed) = 0.3
16. Distribution in Triton supernatant and cytoskeletons	Mainly supernatant

lished in extenso: these correspond to the *E. coli* K-12 protein-gene database (14, 23) and to the rat REF52 database (18, 22).

The *E. coli* K-12 cellular protein-gene database is perhaps the most complete of all databases reported so far and eventually it should trace each protein back to its structural gene. Information contained in this database includes: gene/protein name (protein name, EC number, gene name); 2-dimensional gel spot designations (x-y coordinates from reference gels, alphanumeric designation); genetic information (linkage map location, physical map location, Genebank code, sequence reference, location on Kohara clones); biochemical information (molecular weight, pI, number of residues of each amino acid, mole percent of each amino acid, total number of amino acids in a polypeptide), and regulatory information (cellular level of protein in different media and different temperature, member of regulon, member of stimulon). Major advances of this database are envisaged in the future in view of the eminent sequencing of

the whole *E. coli* genome as well as the development of improved methods to express cloned genes.

The rat REF52 2-dimensional gel protein database lists about 1600 proteins that have been recorded using the QUEST analysis system (18, 22). Included in this quantitative database are 1) protein names (cytoskeletal and heat shock proteins as well as various nuclear, mitochondrial, and cytoplasmic proteins), 2) annotations (subcellular localization, modification, recognition by specific antibodies, coprecipitation, NH₂-terminal sequence, cross-reference to protein sequence information and references to the literature), 3) protein sets (cytoskeletal proteins, phosphoproteins, sets of proteins with PCNA/cyclin-like properties, etc.) and 4) general quantitative data (protein synthesis during growth of normal REF52 cells to confluence and quiescence, and after restimulation of growth-inhibited cells).

In addition to the 2-dimensional gel databases mentioned so far there are several smaller cellular databases being established in human (normal human diploid fibroblasts, lym-

phocytes, leukocytes, leukemic cells) mouse (NIH/3T3 cells, T lymphocytes), *Apisic*, yeast (*Saccharomyces cerevisiae*), plants (wheat, barley, sorghum), and *Euglena*. Databases of tissue protein, (brain, whole mouse, liver) and body fluid proteins (plasma proteins, cerebrospinal fluid, urine, and milk) are being established in several laboratories. The reader is directed to the review by Celis et al. (4) for details and references concerning these databases.

MICROSEQUENCING HAS ADDED A NEW DIMENSION TO COMPREHENSIVE 2-DIMENSIONAL GEL DATABASES: A DIRECT LINK BETWEEN PROTEINS AND GENES

The development of highly sensitive amino acid gas-phase or liquid-phase sequencers (24), together with the establishment of efficient protein and peptide sample preparation methods, has opened the possibility to perform a systematic sequence analysis of proteins resolved by 2-dimensional gel electrophoresis. Indeed, generated pieces of protein sequences can be used to search for protein identity (comparison with available sequences stored in databanks) as well as for preparing specific DNA probes for cloning of as yet uncharacterized proteins (Fig. 1). In addition, partial protein sequences can be stored in 2-dimensional gel databases (for example, see Fig. 2H) and offer a unique link between proteins and genes (Fig. 1).

In the early 1970s gel electrophoresis was used to purify proteins for sequencing purposes (reviewed by Weber and Osborn in ref 25). Proteins were recovered by diffusion and sequenced by the manual dansyl-Edman degradation at the nanomole level. This technique was further refined by using electro-elution to recover proteins and by miniaturizing the system (26). This method has been used extensively, but showed increasing drawbacks (low yields, protein samples contaminated by free amino acids, and NH_2 -terminal blocking) as the amounts of handled protein gradually became smaller (e.g., at the 10 picomol level).

Most of the problems referred to above have been minimized with the introduction of protein-electroblotting procedures (27-32). When proteins are blotted on chemically inert membranes, it is possible to sequence the immobilized proteins directly without additional manipulations. Thus, depending on the amount of bound protein and its nature, this direct sequencing procedure generally yields NH_2 -terminal sequences containing 10-40 residues. As such, this technique was used to identify, by their NH_2 -terminal sequences, differentially expressed major proteins from total cellular extracts separated on 2-dimensional gels. A major difficulty encountered in this procedure is the occurrence of frequent artefactual blockage of the proteins. Several studies suggest that this phenomenon is mainly due to reaction with contaminants (particularly unpolymerized acrylamide present in the gel) and to a high dilution of the protein (low concentration of the protein per unit membrane surface). In addition to this primarily technical problem, many proteins are blocked in vivo by acylation or by a pyrrolidone carboxylic acid cap.

The problem of partial or complete NH_2 -terminal blockage can be circumvented by generating internal amino acid sequences. This is achieved by fragmenting the protein present in the gel (gel in situ cleavage) or by cleaving it while bound to the membrane (membrane in situ cleavage) (33-35). In both cases, proteins are either cleaved in a restricted way (e.g., by limited enzymatic digestion or by using restriction chemical cleavage conditions) or fragmented into smaller peptides.

Of the different combinations examined, we had good results by using exhaustive proteolytic digestion of membrane-immobilized proteins. This method has been described for Ponceau red-stained proteins on nitrocellulose blots (34), for Amido-black-stained Immobilon-bound proteins, and for fluorescamine-detected proteins on glass fiber membranes (35). The proteases used (trypsin, chymotrypsin, or pepsin) cleave at multiple sites, generating small peptides that elute from the blot into the digestion buffer from which they are purified by reversed-phase high performance liquid chromatography (HPLC) before being sequenced individually. Although each of these manipulations could be expected to result in a reduced yield of final sequence information, we were surprised that the peptides could be sequenced with high efficiency. In our hands, this approach could be routinely applied to gel-purified proteins available in amounts ranging from 5 to 10 μg , and often yielded sequence information covering more than 30% of the total protein. As membrane-immobilized proteins are not homogeneously digested, but rather show protease sensitivity next to resistant regions, the number of peptides generated is much lower than expected from the number of potential cleavage sites. Consequently, HPLC peptide chromatograms are less complex and most peptides can be recovered in pure form.

As only limited amounts of a protein mixture can be loaded on a 2-dimensional gel, proteins of interest are often obtained in yields insufficient for the currently available sequencing technology. More material can be obtained by enriching for a certain subcellular fraction (purified cell organelles) or by exploiting affinity (dyes, metals, drugs, etc) or hydrophobic properties of proteins before gel analysis. All of the sequencing results accumulated so far in the human protein database (20) (a few are shown in Fig. 2H) have been obtained from analysis of protein spots collected from 2-dimensional gels that had been stained with Coomassie blue according to standard procedures and dried for storage. Proteins are recovered from the collected gel pieces by a protein-elution-concentration device, combined with gel electrophoresis and electroblotting. Details of this technique have been reported in a previous communication (42) and a brief outline is given below.

Combined gel pieces are allowed to swell in gel sample buffer (a total volume of 1.5 ml). The gel pieces combined with the supernatant are then collected into a large slot made in a new gel. The slot is further filled with Sephadex G-10 equilibrated in gel sample buffer. During consecutive gel electrophoresis, most of the electrical current passes on the side of the slot instead of passing through the slot. This results in both a vertical stacking and horizontal contraction of the protein band. With this device the protein is efficiently eluted from the gel pieces and concentrated from a large volume into a narrow spot. The highly concentrated (about 5 mm^2) protein spot is then electroblotted on PVDF-membranes, stained with Amido black, and in situ digested with trypsin. The peptides generated during digestion elute from the membrane into the supernatant, and can be separated by narrow bore reversed-phase HPLC and collected individually for sequence analysis.

Using this and previous procedures (37, 39, 42), we have so far analyzed 70 protein spots collected from 2-dimensional gels (20, and unpublished observations) (see for example Fig. 2H). The sequence information amounts to 2100 allocated residues corresponding to an average of 30 residues per protein spot. So far we have made cDNAs of many of the unknown proteins that have been microsequenced, and a substantial number has been cloned and sequenced. All available information indicates that it may be possible to obtain partial sequence information from most of

the proteins that can be visualized by Coomassie Brilliant Blue staining.

Partial protein sequences are stored in the database as displayed in Fig. 2H, and it should be possible in the near future to interface this information with forthcoming DNA sequence data from the human genome project. In the long run, as the human genome sequences become available it will be possible to assign partial protein sequences to genes for which the full DNA sequence and chromosomal location are known (Fig. 1).

SUMMARY

The studies presented in this brief review are intended to demonstrate the usefulness of computer-aided 2-dimensional gel electrophoresis and microsequencing to analyze cellular protein patterns, and to link protein and DNA information. As more information is gathered worldwide, comprehensive databases will depict an integrated picture of the expression levels and properties of the thousands of proteins that orchestrate most cellular functions.

Clearly, databases allow easy access to a large body of data and provide an efficient medium to communicate standardized protein information. In the future, databases will foster a wide variety of biological information that can be used to support collaborative research projects in basic and applied biology as well as in clinical research (2, 5, 46). Once a protein is identified in a particular database all the information gathered on it can be made available to the scientist. However, many problems must be solved before protein databases become of general use to the scientific community. A most urgent one is to promote standardization of the gel running conditions so that data produced in a given laboratory may be used worldwide. Surprisingly, the gel running technology as it stands today is still a craftsmanship art.

Finally, comprehensive, computerized databases of proteins, together with recently developed techniques to microsequence proteins, offer a new dimension to the study of genome organization and function (Fig. 1). In particular, human protein databases may become increasingly important in view of the concerted effort to map and sequence the entire human genome. This formidable task is expected to dominate biological research in the next decades. [FJ]

We would like to thank S. Himmelstrup Jørgensen for typing the manuscript and O. Sønderkov for photography. Work in the authors' laboratories was supported by grants from the Danish Biotechnology Programme, the Danish Cancer Foundation, and the Commission of the European Communities.

REFERENCES

- O'Farrell, P. H. (1975) High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **250**, 4007-4021
- Special Issue: Two-dimensional gel electrophoresis. *Clin. Chem.* **28**, 1982
- Celis, J. E., and Bravo, R., eds. (1984) *Two-Dimensional Gel Electrophoresis of Proteins: Methods and Applications*. Academic, New York
- Celis, J. E., Madsen, P., Gesser, B., Kwee, S., Nielsen, H. V., Rasmussen, H. H., Honoré, B., Leffers, H., Ratz, G. P., Basse, B., Lauridsen, J. B., and Celis, A. (1989) Protein databases derived from the analysis of two-dimensional gels. In *Advances in Electrophoresis* (Chrambach, C., ed) VCH, Weinheim, Germany
- Special Issue: Two-dimensional gel electrophoresis in cell biology. (Celis, J. E., ed) *Electrophoresis* **11**, 1990
- Celis, J. E., Honoré, B., Bauw, G., and Vandekerckhove, J. (1990) Comprehensive computerized 2D gel protein databases offer a global approach to the study of the mammalian cell. *BioEssays* **12**, 93-98
- Garrels, J. I. (1983) Two-dimensional gel electrophoresis and computer analysis of proteins synthesized by cloned cell lines. *Methods Enzymol.* **100**, 411-423
- Anderson, N. L., Hofmann, J. P., Gemmel, A., and Taylor, S. (1984) Global approaches to the quantitative analysis of gene expression patterns observed by two-dimensional gel electrophoresis. *Clin. Chem.* **30**, 2031-2036
- Garrels, J. I., Farrar, J. T., and Burwell, C. B. (1984) The Quest system for computer-analyzed two-dimensional electrophoresis of proteins in *Two-Dimensional Gel Electrophoresis of Proteins: Methods and Applications* (Celis, J. E., and Bravo, R., eds) pp 37-91. Academic, New York
- Vincens, P., and Tarroux, P. (1988) Two-dimensional electrophoresis computerized processing. *Int. J. Biochem.* **20**, 499-509
- Appel, R., Hochstrasser, D., Roch, C., Funk, M., Müller, A. F., and Pellegrini, C. (1988) Automatic classification of two-dimensional gel electrophoresis pictures by heuristic clustering analysis: a step toward machine learning. *Electrophoresis* **9**, 136-142
- Lemkin, P. F., and Lester, E. P. (1989) Database and search techniques for two-dimensional gel protein data: a comparison of paradigms for exploratory data analysis and prospects for biological modeling. *Electrophoresis* **10**, 122-140
- Miller, M. J. (1989) Computer-assisted analysis of two-dimensional gel electrophoretograms. *Adv. Electrophoresis* **3**, 182-217
- Phillips, T. D., Vaughn, V., Bloch, P. L., and Neidhardt, F. C. (1987) In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology. Gene-Protein Index of Escherichia coli K-12*, 2 ed. (Neidhardt, F. C., Ingraham, J. L., Low, K. B., Magasanik, B., Schaechter, M., and Umberger, H. E. ed) pp. 919-966. American Society for Microbiology, Washington, DC.
- Celis, J. E., Ratz, G. P., Celis, A., Madsen, P., Gesser, B., Kwee, S., Madsen, P. S., Nielsen, H. V., Yde, H., Lauridsen, J. B., and Basse, B. (1988) Towards establishing comprehensive databases of cellular proteins from transformed human epithelial amnion cells (AMA) and normal peripheral blood mononuclear cells. *Leukemia* **9**, 561-601
- Special Issue: Protein databases in two-dimensional electrophoresis. (Celis, J. E., ed) *Electrophoresis* **2**, 1989
- Celis, J. E., Ratz, G. P., Madsen, P., Gesser, B., Lauridsen, J. B., Brogaard-Hansen, K. P., Kwee, S., Rasmussen, H. H., Nielsen, H. V., Crüger, D., Basse, B., Leffers, H., Honoré, B., Møller, O., and Celis, A. (1989) Computerized, comprehensive databases of cellular and secreted proteins from normal human embryonic lung MRC-5 fibroblasts: identification of transformation and/or proliferation sensitive proteins. *Electrophoresis* **10**, 76-115
- Garrels, J. I., and Franza, B. R. (1989) The REF52 protein database. Methods of database construction and analysis using the Quest system and characterizations of protein patterns from proliferating and quiescent REF52 cells. *J. Biol. Chem.* **264**, 5283-5298
- Celis, J. E., Crüger, D., Kiil, J., Dejgaard, K., Lauridsen, J. B., Ratz, G. P., Basse, B., Celis, A., Rasmussen, H. H., Bauw, G., and Vandekerckhove, J. (1990) A two-dimensional gel protein database of noncultured total normal human epidermal keratinocytes: identification of proteins strongly up-regulated in psoriatic epidermis. *Electrophoresis* **11**, 242-254
- Celis, J. E., Gesser, B., Rasmussen, H. H., Madsen, P., Leffers, H., Dejgaard, K., Honoré, B., Olsen, E., Ratz, G., Lauridsen, J. B., Basse, B., Mouritzen, S., Hellerup, M., Andersen, A., Walbum, E., Celis, A., Bauw, G., Puype, M., Van Damme, J., and Vandekerckhove, J. (1990) Comprehensive two-dimensional gel protein databases offer a global approach to the analysis of human cells: the transformed amnion cells (AMA) master database and its link to genome DNA sequence data. *Electrophoresis* **12**, 989-1071

21. Celis, J. E., Dejgaard, K., Madsen, P., Leffers, H., Gesser, B., Honoré, B., Rasmussen, H. H., Olsen, E., Lauridsen, J. B., Ratz, G., Mouritzen, S., Hellerup, M., Andersen, A., Walbum, E., Celis, A., Bauw, G., Puype, M., Van Damme, J., and Vandekerckhove, J. (1990) The MRC-5 human embryonal lung fibroblast two-dimensional gel cellular protein database: quantitative identification of polypeptides whose relative abundance differs between quiescent, proliferating and SV40 transformed cells. *Electrophoresis* 12, 1072-1113.
22. Garrels, J. I., Franza, B. R., Chang, C., and Latter, G. (1990) Quantitative exploration of the REF52 protein database: cluster analysis reveals the major protein expression profiles in responses to growth regulation, serum stimulation, and viral transformation. *Electrophoresis* 12, 1114-1130.
23. Van Bogelen, R. A., Hutton, M. E., and Neidhardt, F. C. (1990) Gene-protein database of *Escherichia coli* K-12, 3rd ed. *Electrophoresis* 12, 1131-1166.
24. Hewick, R. M., Hunkapiller, M. W., Hood, L. E., and Dreyer, W. J. (1981) A gas-liquid solid phase peptide and protein sequencer. *J. Biol. Chem.* 256, 7990-7997.
25. Weber, K., and Osborn, M. (1985) In *The Proteins and Sodium Dodecyl Sulfate: Molecular Weight Determination on Polyacrylamide Gels and Related Procedures* (Neurath, H. et al., eds) Vol. 1, pp. 179-223. Academic, New York.
26. Hunkapiller, M. W., Lujan, E., Ostrander, F., and Hood, L. E. (1983) Isolation of microgram quantities of proteins from polyacrylamide gels for amino acid sequence analysis. *Methods Enzymol.* 91, 227-236.
27. Vandekerckhove, J., Bauw, G., Puype, M., Van Damme, J., and Van Montagu, M. (1985) Protein-blotting on polybrene-coated glass-fiber sheets. *Eur. J. Biochem.* 152, 9-19.
28. Aebersold, R. H., Teplov, D. B., Hood, L. E., and Kent, S. B. H. (1986) Electroblooming onto activated glass. *J. Biol. Chem.* 261, 4224-4238.
29. Bauw, G., De Loose, M., Inzé, D., Van Montagu, M., and Vandekerckhove, J. (1987) Alterations in the phenotype of plant cells studied by NH₂-terminal amino acid-sequence analysis of proteins electroblotted from two-dimensional gel-separated total extracts. *Proc. Natl. Acad. Sci. USA* 84, 4806-4810.
30. Matsudaira, P. (1987) Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *J. Biol. Chem.* 262, 10035-10038.
31. Eckerskorn, C., Mewes, W., Goretzki, H., and Lottspeich, F. (1985) A new siliconized-glass fiber as support for protein-chemical analysis of electroblotted proteins. *Eur. J. Biochem.* 176, 509-519.
32. Moos, M., Jr., Neuvén, N. Y., and Liu, T.-Y. (1988) Reproducible high yield sequencing of proteins electrophoretically separated and transferred to an inert support. *J. Biol. Chem.* 263, 6005-6008.
33. Kennedy, T. E., Gawinowicz, M. A., Barzilai, A., Kandel, E. R., and Sweatt, J. D. (1988) Sequencing of proteins from two-dimensional gels by using in situ digestion and transfer of peptides to polyvinylidene difluoride membranes: application to protein associated with sensitization in *Aplysia*. *Proc. Natl. Acad. Sci. USA* 85, 7008-7012.
34. Aebersold, R. H., Leavitt, J., Saavedra, R. A., Hood, L. E., and Kent, S. B. H. (1987) Internal amino acid sequence analysis of protein separated by one- or two-dimensional gel electrophoresis after in situ protease digestion on nitrocellulose. *Proc. Natl. Acad. Sci. USA* 84, 6970-6972.
35. Bauw, G., Van Den Bulcke, M., Van Damme, J., Puype, M., Van Montagu, M., and Vandekerckhove, J. (1988) Protein electroblotting on polybase-coated glassfiber and polyvinylidene difluoride membranes: an evaluation. *J. Prot. Chem.* 7, 194-200.
36. Celis, J. E., Ratz, G. P., Madsen, P., Gesser, B., Lauridsen, J. B., Leffers, H., Rasmussen, H. H., Nielsen, H. V., Cruiger, D., Basse, B., Honoré, B., Møller, O., Celis, A., Vandekerckhove, J., Bauw, G., Van Damme, J., Puype, M., and Van Den Bulcke, M. (1989) Comprehensive human cellular protein databases and their implication for the study of genome organization and function. *FEBS Lett.* 244, 247-254.
37. Bauw, G., Van Damme, J., Puype, M., Vandekerckhove, J., Gesser, B., Lauridsen, J. B., Ratz, G. P., and Celis, J. E. (1989) Protein-electroblotting and -microsequencing strategies in generating protein databases from two-dimensional gels. *Proc. Natl. Acad. Sci. USA* 86, 7701-7705.
38. Aebersold, R., and Leavitt, J. (1990) Sequence analysis of proteins separated by polyacrylamide gel electrophoresis: Towards an integrated protein database. *Electrophoresis* 11, 517-527.
39. Bauw, G., Rasmussen, H. H., Van Den Bulcke, M., Van Damme, J., Puype, M., Gesser, B., Celis, J. E., and Vandekerckhove, J. (1990) Two-dimensional gel electrophoresis, protein electroblotting and microsequencing: a direct link between proteins and genes. *Electrophoresis* 11, 528-536.
40. Tempst, P., Link, A. J., Riviere, L. R., Fleming, M., and Elion, C. (1990) Internal sequence analysis of protein separated on polyacrylamide gels at the submicrogram level: improved methods, applications and gene cloning strategies. *Electrophoresis* 11, 537-553.
41. Eckerskorn, C., and Lottspeich, F. (1990) Combination of two-dimensional gel electrophoresis with microsequence and amino acid composition analysis: improvement of speed and sensitivity in protein characterization. *Electrophoresis* 11, 554-561.
42. Rasmussen, H. H., Van Damme, J., Bauw, G., Puype, M., Gesser, B., Celis, J. E., and Vandekerckhove, J. (1991) In *Methods in Protein Sequence Analysis* (Jörnvall, H., and Höög, J. O., eds) pp. 103-114. Eighth International Conference on Methods in Protein Sequence Analysis, Birkhäuser Verlag, Boston.
43. Olson, A. D., and Miller, M. J. (1988) Elsie 4: quantitative computer analysis of sets of two-dimensional gel electrophoretograms. *Anal. Biochem.* 169, 49-70.
44. Vincens, P., Paris, N., Pujol, J. L., Gaboriaud, C., Rabilloud, T., Penetier, J., Matherat, P., and Tarroux, P. (1986) HERMES: a second generation approach to the automatic analysis of two-dimensional electrophoresis gels. Part I: Data acquisition. *Electrophoresis* 7, 347-356.
45. Celis, J. E., Madsen, P., Celis, A., Nielsen, H. V., and Gesser, B. (1987) Cyclin (PCNA, auxiliary protein of DNA polymerase- δ) is a central component of the pathway(s) leading to DNA replication and cell division. *FEBS Lett.* 220, 1-7.
46. Anderson, N. G., and Anderson, N. L. (1982) The human protein index. *Clin. Chem.* 28, 734-748.

THIS PAGE BLANK (USPTO)

Bo Franzén¹
 Stig Linder¹
 Ken Okuzawa²
 Harabumi Kato²
 Gert Auer¹

¹Division of Tumor Pathology,
 Department of Pathology, Division
 of Experimental Oncology,
 Karolinska Hospital and Institute,
 Stockholm Sweden

²Tokyo Medical College, Department
 of Surgery, Tokyo

³Division of Experimental Oncology,
 Karolinska Hospital and Institute,
 Stockholm

Nonenzymatic extraction of cells from clinical tumor material for analysis of gene expression by two-dimensional polyacrylamide gel electrophoresis

We have compared different methods of preparation of malignant cells for two-dimensional electrophoresis (2-DE). We found all methods using fresh tissue to be superior compared to methods using frozen tissue. Our results indicate that nonenzymatic methods of preparation of tumor cells, including fine needle aspiration, scraping and squeezing, have advantages over methods using enzymatic extraction of cells. Nonenzymatic methods are rapid, appear to reduce loss of high molecular protein species, and alleviate the necessity of separating viable and nonviable cells by Percoll gradient centrifugation. Using these techniques, high-quality 2-DE maps were derived from tumors of the lung and breast. In the resulting polypeptide patterns, heat shock proteins, non-muscle tropomyosins and intermediate filament were identified. We conclude that nonenzymatic extraction of malignant cells from fresh tumor tissue improves the possibilities that these techniques may be useful in clinical diagnosis.

1 Introduction

Tumors may develop by a number of different mechanisms in any given cell type. At the time of diagnosis, tumors will have progressed along different pathways to various stages of malignancy. To provide a basis for individual therapy it is of importance to examine specific properties of the tumor cell population in each patient. A large number of different markers have been described in order to increase the diagnostic accuracy. It is likely that a combination of several markers is needed in the future in order to reflect different properties of the tumor. One important method for the resolution of a large number of potential markers is two-dimensional electrophoresis (2-DE). Extensive efforts are being made in identifying various polypeptides separated by 2-DE and to characterize how the expression of these polypeptides is affected by the response to cellular transformation and various culture conditions [1,2]. It would be of value to transfer this information to 2-DE separations of polypeptides from tumor tissue samples. However, one prerequisite is that the quality of the 2-DE gels from tumor samples is comparable in quality with 2-DE gels from samples of cultured cells.

Frozen tumor tissues are commonly used for various biochemical assessments. However, if such samples are analyzed by 2-D polyacrylamide gel electrophoresis (PAGE), the polypeptide patterns are obscured by contamination of serum- and connective tissue proteins. Such nontumor-cell-related variations represent serious problems in the interpretation and inter-patient comparison of 2-DE

patterns [3]. 2-DE patterns of cells prepared from fresh tumor material were analyzed after enzymatic extraction of tumor cells [4, 5] or after culturing tumor fragments in medium containing radioactive amino acids [6]. These procedures may, however, lead to alterations in the gene expression/polypeptide patterns. We are only aware of one study where nonenzymatic extraction of cells from fresh tumor tissue (prostate cancer) was used to prepare samples for 2-D PAGE [4]. We have examined enzymatic extraction and various nonenzymatic preparation techniques, including fine needle aspiration, for the preparation of cells from fresh tumor tissues. We describe nonenzymatic extraction procedures that are rapid, lead to high-quality 2-DE patterns, and that alleviate the necessity to purify tumor cell populations from dead cells.

2 Materials and methods

2.1 Cell cultures and samples used for spot identification

A rat embryonal fibroblast cell line, WT2 (a kind gift from Dr. J. I. Garrels and Dr. S. Pattersson) was used for the identification of a number of heat shock and structural proteins. Human normal diploid lung fibroblasts, WI38, human epithelial breast carcinoma cells, MDA-231 and MCF-7 were purchased from ATCC and grown as recommended. Polypeptides prepared from a leukemia type pre-B-ALL were separated by 2-DE. The 2-DE map was then analyzed by Dr. S. M. Hanash (University of Michigan, Ann Arbor, USA).

2.2 Tumor tissues samples

In this study, 2-DE maps from seven tumors were used as representative illustrations: two adenocarcinoma of the lung (LA, and LB, mucinous, both cases intermediate grade of differentiation), one squamous carcinoma of the lung (LS), one carcinoid-like breast cancer (BC), one microfollicular adenoma (highly differentiated) of the thyroid (TA), one highly differentiated hyperneph-

Correspondence: Dr. Bo Franzén, Division of Tumor Pathology, Department of Pathology, L1:01, Karolinska Hospital and Institute, 10401 Stockholm 60, Sweden

Abbreviations: 2-DE, Two-dimensional polyacrylamide gel electrophoresis; IEF, isoelectric focusing; LDH, lactate dehydrogenase; NP-40, Nonidet P-40; PBS, phosphate buffered saline; PCNA, proliferating cell nuclear antigen; PIH, protease inhibitors; PMSF, phenylmethyl sulfonyl fluoride; SDS, sodium dodecyl sulfate; WW, wet weight

roma, a tumor of the kidney (KH), and finally one case of poorly differentiated corpus carcinoma (CP).

2.3 Preparation of cultured cells

The cell monolayers were washed twice in phosphate buffered saline (PBS) and then scraped off in ice-cold PBS including protease inhibitors (PIH), phenylmethylsulfonyl fluoride (PMSF) 0.2 mM and 0.83 mM benzamidine pelleted at $660 \times g$, 3 min (-4°C) and washed one time before final centrifugation at $2700 \times g$, 5 min. The wet weight of the cell pellet was recorded and the cells were stored at -80°C until further processing.

2.4 Preparation of tumor tissue samples

2.4.1 General remarks

Macroscopically representative and non-necrotic tumor tissues were selected within 20 min after resection. Parallel samples were routinely prepared for cytology. The samples were processed as rapidly as possible on ice or at -4°C and in the presence of PIH. Cells were stained with DiffQuick (Baxter) and usually examined at three different occasions during the preparation procedure: (i) cytology sample, (ii) extracted cells and (iii) cells after percoll gradient centrifugation.

2.4.2 Specimen acquisition

The strategy of sample preparation is shown in Fig. 1. Tumor tissue cell samples were usually obtained by fine needle aspiration (NA) using a 0.7 mm needle. The syringe was filled with 1–2 mL of ice-cold culture medium/PIH. We found that if a tumor appeared to be very fibrous it is difficult to extract enough cells for 2-DE analysis. In these cases, two alternative techniques were examined. (i) The tumor was cut in the middle and the fresh surface scraped (SC) by a scalpel. The cell-rich material was then transferred to ice-cold culture medium (L15 with 5% fetal calf serum)/PIH. (ii) A part of the tumor sample was placed in culture medium on ice for further processing at the laboratory in the following way: the material was cut into very small fragments on a pre-cooled dissection plate and transferred to a small glass chamber with a 0.7 mm metal net 5 mm above the bottom of the chamber. Medium/PIH was added to cover the sample (8 mL) which was gently squeezed (SQ) towards the net in order to release and wash out cells. NA and SC were also compared with an enzymatic extraction (EE) procedure described previously [5]. Briefly, thin slices of tissue were incubated with collagenase (1 mg/mL) and elastase (2 mg/mL) in medium for 1 h at 37°C . Extracted cells from every sample were then subjected to percoll gradient centrifugation (Section 3.2.3).

2.4.3 Separation of cells by Percoll gradient centrifugation

The cell suspension was filtered through two nylon mesh filters, (i) 250 μm and (ii) 100 μm and then centrifuged

at $660 \times g$ for 3 min. The cell pellet was resuspended carefully in medium, using a syringe and loaded onto a two-step discontinuous Percoll/PBS gradient: 20.4% (density = 1.03 g/mL) and 54.7% (density = 1.07 g/mL) and centrifuged at $1000 \times g$ for 15 min. In this system, dead cells stay on the top, viable cells sediment to the interphase and erythrocytes sediment to the bottom. The viability of cells in the top fraction and interphase was checked by the trypan blue exclusion test. The interphase cell layer ($> 90\%$ viability) was collected and washed one time in a large volume PBS/PIH (centrifuged at $800 \times g$ for 3 min). Finally, the cells were resuspended in 1.4 mL PBS and pelleted at $2700 \times g$ for 5 min. The wet weight (WW) was recorded and the pellet was then stored at -80°C .

2.4.4 Final preparation of cells for 2-D PAGE analysis

From this point, cultured cell samples were treated in the same way as tumor cell samples. Each cell pellet was thawed on ice and resuspended in 1.89 μL mQ water per mg WW ($= 1.89 \times \text{WW}$) μL . The suspension was frozen and thawed 4–5 \times to break the cells [7]. A volume of $(0.089 \times \text{WW})$ μL 10% sodium dodecyl sulfate (SDS), including 33.3% mercaptoethanol, was mixed with the sample and incubated 5 min on ice with $(0.329 \times \text{WW})$ μL of a solution of DNase I (0.144 mg/mL 20 mM Tris-HCl with 2 mM $\text{CaCl}_2 \times 2\text{H}_2\text{O}$, pH 8.8) and RNase A (0.0718 mg/mL Tris) [8,9]. The sample was frozen and lyophilized. Sample buffer [10] including

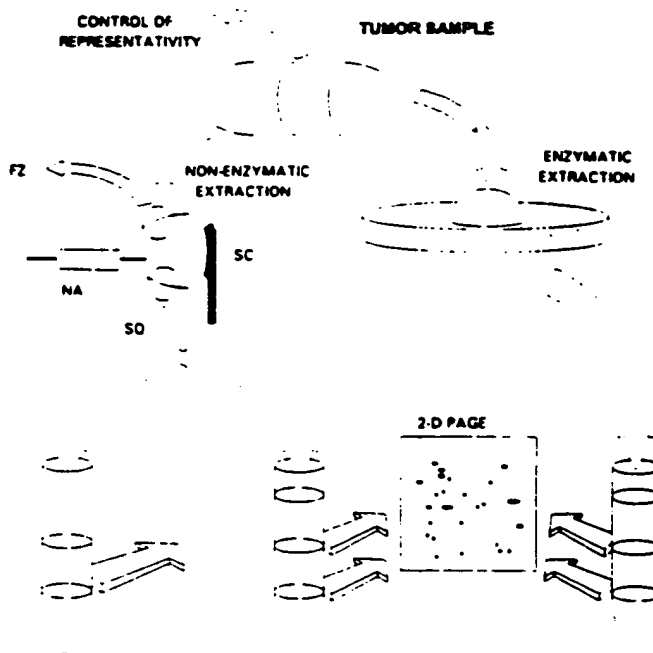


Figure 1 Experimental flow chart showing main steps of the preparation procedures. The abbreviations used for nonenzymatic extraction procedures are: FZ: frozen sample preparation; NA: needle aspiration; SC: scraped; and SQ: squeezed sample. Extracted cells are then loaded as a suspension (top volume of each tube) onto either 1.07 g/mL Percoll (left), or a discontinuous Percoll gradient from the nonenzymatic extraction (middle), or from enzymatic extraction (right). Cellular top- and interphase fractions are then used for 2-DE. For details see Section 2.